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 46 No. 12 December 1976

The Radio and Electronic Engineer

The Journal of the Institution of Electronic and Radio Engineers

Value and Communication

AT this time every year a process is set in action, initially within the Institution's Papers Committee, which has the object of identifying the outstanding contributions from among the seventy or so papers published in *The Radio and Electronic Engineer* during the past twelve months. The recommendations go to the Council for consideration and an announcement is made during the following summer of the Premiums that are to be presented by the President at the Annual General Meeting in the autumn. The Institution has awarded Premiums ever since 1946: originally there were just four Premiums—Clerk Maxwell, Heinrich Hertz, Marconi and Partridge—but with the growing scope of radio and electronic engineering additional specialized awards have been established. Their award represents a cachet of approval which we know is much valued by the successful authors, quite apart from the monetary value.

During the past thirty years, criteria have been evolved for deciding just what makes an outstanding paper and, by assessing the extent to which particular papers meet these criteria, using a simple marking system, an order of merit can be set out. Thus, from the initial seventy papers, perhaps a third will have already been shortlisted for detailed examination by members of the Papers Committee and the Group Committees and as a result of the marks given usually about a dozen papers emerge as outstanding and are recommended for Premiums relevant to their subjects. Frequently a Premium has to be withheld because among the papers of merit one has not been published falling within the terms of award.

This careful and rather lengthy process clearly depends for its validity on the initial criteria and these are currently, in order of priority: originality, importance, and presentation. It is thought that most Institutions and societies use a scheme of this kind to carry out a difficult task. Some time ago, a distinguished mechanical engineer, Dr. K. J. Durrands, Rector of Huddersfield Polytechnic, advocated criteria which, though broadly similar to those of the IERE, are phrased to emphasize engineering merit, namely: intellectual content, value to industry, and how quickly and easily the information in the paper can be used by the practising engineer.

Although these are criteria developed primarily for assessing papers for awards, considerations along these lines can readily be applied in deciding on the suitability of papers for publication.

There is however a deduction to be made by the perceptive intending author on reading these criteria. While he will appreciate that award of a Premium is more likely to result from bearing them in mind, they represent, particularly as set out by Dr. Durrands, concepts which ought to be continually before every engineer who writes a paper for his professional journal. A keener appreciation of the ultimate value of the work described and the rapid and easy use of information presented must lead to the eventual emergence of a class of paper that will appeal to a wide readership, without lowering sights to easy popularization. 'Is my paper really necessary?' could be one question an author might ask, followed by the equally self-critical one of 'Am I putting my subject over in the most effective way?'.

The first and most obvious purpose of Institution Premiums is to reward achievement and by so doing to encourage other engineers to describe their work. But a more important factor is in setting standards so that it is apparent what is expected of a 'good' paper. Most engineers will agree that the outward looking attributes of value and communication in the broadest sense are just as necessary as originality or intellectual content.

F.W.S.

Contributors to this issue*

Dennis Ralphs (Member 1969, Associate 1945) graduated from Birmingham University in 1943. and during the war worked on the development of airborne radar and ground training simulators, and subsequently on instrumentation for nuclear research. In 1947 he joined the Scientific Civil Service and since 1954 has been head of the Research and Development section of the Foreign and Commonwealth Office Communications Engineering Department. He was responsible for the design of the 'Piccolo' 32-tone telegraphy system used on the FCO h.f. radio network and has continued the study of m.f.s.k. principles and the practical application of these to poor-signal telegraphy.

Francis Sladen was employed by the Communications Engineering Department of the British Foreign and Commonwealth Office during the period 1967 to 1974, and is currently working as a research fellow in the Optical Fibre Communications Group at Southampton University. As an undergraduate he read electronics at Southampton University, graduating in July 1973 with a first class honours degree in electronic engineering. Recent research has been in the field of leaky modes on graded index fibres, on the development of a method for measuring the refractive index profile in optical fibres and currently, on optical fibre communication systems.

Raymond Whittle did his National Service in the Royal Air Force as a radar mechanic and in 1951 joined Mullard where he was engaged in quality control and later in applications engineering of gas tubes. During this period he studied part-time at Battersea Polytechnic for the B.Sc. degree of London University in physics and mathematics. Mr. Whittle joined Standard Telephones & Cables Ltd. (ITT) in 1956 as an Applications Engineer and gained diverse experience on power and microwave tubes; later he was concerned with solid-state devices and publicity activities. From 1963 to 1975 he was Head of the Applications Engineering Section of the Valve Division at Paignton and he is currently engaged in market research in the Electron Device Division. He has contributed several articles to the Technical Press on electronic tubes and semiconductor devices.

John Behenna received the Diploma in Electrical Engineering from Faraday House Engineering College in 1952. He had spent three years in the Radio Branch of the Royal Navy before going to college. He then joined Standard Telephone & Cables, New Southgate, to work on the development of high power transmitters and in 1957 transferred to the Valve Division at Paignton where he was in charge of the high-power vacuum-tube development laboratory and other special projects until May of last year. He is now working outside the electronics industry.

Alan Holt (Fellow 1974, Member 1959, Graduate 1953) holds a Personal Chair in Electrical Engineering in the University of Newcastle upon Tyne. He has worked with the Post Office Engineering Department and served with the RAF and following research in the Department of Electronics at the University of Southampton he was awarded the Ph.D. degree in 1959. He went to Newcastle as a Lecturer in 1957 and was later appointed to a Readership. Professor Holt's main teaching and research interests are in circuit and systems studies; computer aided design; computer simulation of systems; and the theory and construction of thin film devices.

John Attikiouzel received the B.Sc. degree in Electrical Engineering from Newcastle University in June 1969. For the next three years he carried out research into computer aided design for active networks and was awarded a Ph.D. in May 1973. After completing his doctoral research, in October 1972 he was appointed as a Research Officer at the Electrical and Electronic Engineering Department at Newcastle University, working on control theory, distributed and active network theory. Since 1975 he has been on the staff of Portsmouth Polytechnic.

Richard Bennett obtained a Ministry of Defence Scholarship to study at the University of Newcastle upon Tyne, where he graduated in electrical engineering in 1972. The following year he received his M.Sc. for a thesis on the design of digital moving target indicators, this work being sponsored by the Royal Radar Establishment, Malvern. A continuation of this study has recently been submitted for his Ph.D. degree.

* See also pages 604 and 616.

UDC 621.391.812.3.029.55

An h.f. channel simulator using a new Rayleigh fading method

J. D. RALPHS, B.Sc., C.Eng., M.I.E.R.E.*

and

F. M. E. SLADEN, B.Sc.†

SUMMARY

The paper describes a new method of producing a discrete-level approximation to a Rayleigh fade by means of a quasi-random digital process. This offers considerable advantages over existing methods in cost, simplicity and flexibility, having an exactly repeatable fading pattern and a fading rate which may be easily changed by varying the digital clock rate. Mathematical analysis and experimental evidence show that the output closely approximates to a Rayleigh fade in all respects. A description is given of a prototype h.f. ionospheric channel simulator incorporating the design. In addition to Riceian fading the prototype allows the simulation of two-path propagation, Doppler shift, and additive Gaussian noise. Some practical applications of the simulator are discussed.

* *Communications Engineering Department, Foreign and Commonwealth Office, Hanslope Park, Hanslope, Milton Keynes MK19 7BH.*

† *Now with the Department of Electronics, University of Southampton.*

1 Introduction

1.1 Concept

The use of real-time analogue simulators to evaluate the effects of ionospheric propagation conditions on h.f. communications systems is well established,[1–3] as it allows the quantitative analysis of the effects of parameters (such as fading rate and signal-to-noise ratio) to an extent which would be totally impractical on a 'live' signal. In the opinion of the authors, current practice in the field of simulator design falls short of the ideal in three respects. Firstly, the mathematical model postulated is often an over-simplification of the effects which can occur on many h.f. channels, secondly, the circuit techniques used do not allow an accurate approximation to the model, and thirdly the range of parameter variation available is often less than the extremes which may be experienced on actual h.f. circuits. Other criticisms are that the simulators are often bulky and expensive, and are difficult and time-consuming to set up and operate, having been designed in most cases as instruments for basic research. The authors consider that there is a need for an h.f. channel simulator design suitable for use in the development and routine testing of communications equipment. Such an instrument will be required to simulate 'typical' or 'extreme' conditions under which an h.f. communication link may be required to operate, calling for the accurate and consistent reproduction of a relatively simple but flexible mathematical model with a wide range of adjustment on each parameter, rather than the simulation of postulated complex ionospheric situations.

1.2 Ionospheric Channel Model

It has been shown[4–6] that many signals on h.f. links show 'Rayleigh fading' characteristics, in that an unmodulated carrier is received with a Rayleigh distribution in amplitude and a Gaussian distribution in power spectrum. A non-fading ('specular') component may also be received, giving a Nakagami-Rice or Riceian amplitude distribution.[4,5,7] Long-term amplitude distributions may approach a log-normal law but this has been ignored for simulation purposes since this is unlikely to have an appreciable influence on equipment design or system performance.

The rapidity of fading is usually expressed in terms of the 'fading rate' which is defined (for a single carrier) as the average number of downward crossings per second of the envelope through the median value. Fading rates of up to 10 fades per second have been reported[8,9] but the authors suggest that simulation of at least 20 fades per second is desirable.

A signal may be received by two or more paths of roughly equal attenuation but differing appreciably in length, causing a time delay T between them and thus creating an 'echo'. If $\frac{1}{2}T$ is less than the baseband of a communication system, such echoes can have a pronounced effect on system performance, and if T is an appreciable fraction of the element length of a telegraphy system, major deterioration in accuracy can result. Differential time delays of up to 3 ms are relatively common on long-distance h.f. circuits and delays of up to 10 ms on a 500 km path have been reported.[11] Practical experience indicates that longer values than this may be encountered, and simulation of differential delays up to about 15 ms is advisable.

The complete frequency spectrum of a signal may be displaced by Doppler effects, and while reliable statistics of a wide variety of paths are not available, frequency shifts of up to 30 Hz have been quoted[8] and it is desirable to simulate shifts up to at least this value.

It is common practice to simulate the reception of thermal and atmospheric noise by the addition of white Gaussian noise to the signal. This choice is dictated not so much by the prevalence of this type of interference on h.f. channels as by its universal acceptance as a standard and the relative simplicity of its mathematical treatment.

These effects may be combined in a single mathematical equation, so that the communication channel is simulated by a four-terminal network which operates on an input signal $V \sin \omega t$ to produce the output signal:

$$V_0 = V\{A_0 \sin \omega t + A_1 B_1(t) \sin [(\omega+\delta\omega)t+\theta_1(t)]\} + + V A_2 B_2(t) \sin [\omega(t+T)+\theta_2(t)] + V_n \quad (1)$$

A_0, A_1, A_2 = fixed attenuations

$B_1(t)$ = time-variable attenuation such that the signal envelope obeys the Rayleigh amplitude distribution

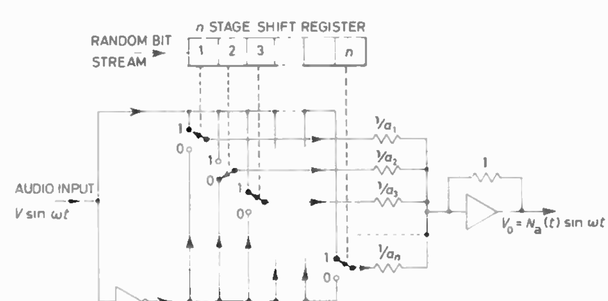
$\theta_1(t)$ = phase variable with random distribution over 0–2π

$B_2(t), \theta_2(t)$ = as $B_1(t), \theta_1(t)$ but uncorrelated with them

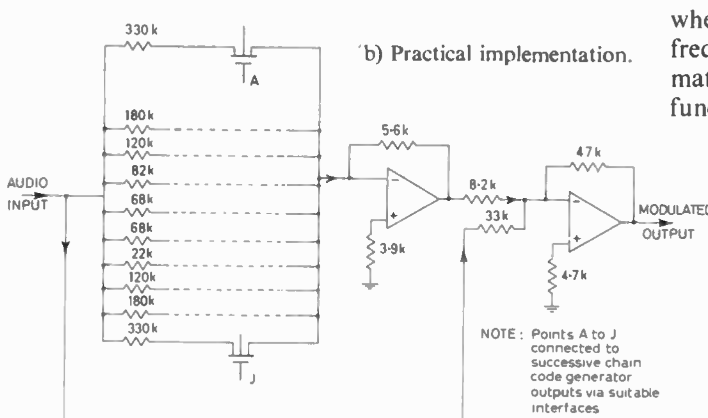
T = multi-path time delay

$\delta\omega$ = Doppler frequency shift

V_n = Gaussian noise voltage.



(a) Basic circuit.



(b) Practical implementation.

Although the separate definitions of B and θ are normally considered sufficient to define a 'Rayleigh fading' function, there must be, implicit in the mechanism of fading, a relationship of some kind between the instantaneous values or time derivatives of the two parameters. This relationship does not seem to be referred to explicitly in the literature but fundamental considerations suggest that the mean rate of change of phase is proportional to the inverse of the absolute amplitude—at least at low amplitudes. This correlation has not been studied in detail in the project discussed here since it was believed that it would be achieved automatically in the method to be described.

In the prototype design the parameters A_0, A_1, A_2, $\delta\omega$, T and V_n are independently controlled variables (except that for simplicity $A_2 = A_1$ or 0), and the fading process may be switched off (making $B(t)$ and $\theta(t)$ into constants). The first- and second-term voltages are independently available at two outputs to allow some facilities for testing diversity systems.

The fading rates of the variables B_1 and B_2 may be varied independently but for operational simplicity it is arranged that the fading period of B_2 is a simple integral multiple (1, 2, 4 etc.) of that of B_1.

Methods of generating white noise voltages, and circuits for frequency translation and time delay are well known and the limitations of performance of such circuits are relatively easy to define and measure. The circuits used in the equipment for these purposes will be described briefly here for the sake of completeness. However, the current methods of generating the Rayleigh fading function are not easily applicable to the type of equipment envisaged (for reasons given below) and the method to be described is believed to be novel and carry considerable advantages. This will therefore be treated in more detail.

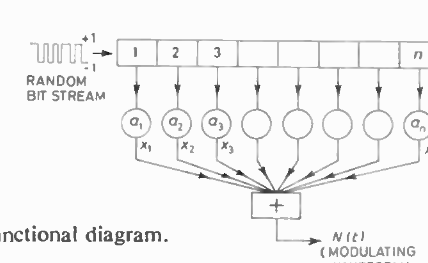
2 Rayleigh Fading Simulation

2.1 Multiple Oscillator System

This is probably the most widely-used method of reproducing the Rayleigh Fading function and operates on the principle of summing several vectors of constant amplitude but varying phase, thus:

$$V(t) = \sum_{m=1}^{M} AV \sin (\omega+\omega_m)t \quad (2)$$

where M is usually 6 or 8 and ω_m is a very small off-tune frequency selected to distribute the M vectors approximately uniformly over a given bandwidth, which is a function of the fading rate required.[1,2]



(c) Functional diagram.

Fig. 1. Gaussian modulator.

The amplitude distribution of such a system is a reasonable approximation to Rayleigh but the power spectrum deviates appreciably from Gaussian. The system is expensive and is time-consuming and inconvenient to use and unless accurately adjusted may result in an undetected specular component in the fading waveform. The range of fading rates reproducible with accuracy is limited.

2.2 Quadrature Noise Modulation

A Rayleigh fade may be generated by phase shifting the input signal to obtain two signal sources in phase quadrature, and subsequently amplitude modulating each quadrature signal source with a low frequency noise source before recombining. The two noise sources must be uncorrelated, and should each have a zero-mean Gaussian amplitude distribution with equal r.m.s. values, and a Gaussian-shaped power spectrum with equal r.m.s. values. The r.m.s. frequency of the power spectra determines the fading rate. The noise generators may be analogue or derived from quasi-random binary streams at high bit rate, the spectrum of the fading signal being determined in either case by low-pass filters, which must be of 3-pole design or better if the spectra are to approach the required Gaussian shape with sufficient accuracy. These filters constitute a major disadvantage of the system, since for each fading rate required, two 3-pole filters must be switched in each fading channel.

2.3 Digital Quadrature Attenuation

The technique to be described may be regarded as a variant of the above, in which the modulating voltages are replaced by variable attenuations at fixed phase thus:

$$V(t) = AN_a(t).V \sin \omega t + AN_b(t).V \sin(\omega t + \tfrac{1}{2}\pi) \qquad (3)$$

where $N_a(t)$ and $N_b(t)$ are discrete random variables (positive and negative), selected from a finite linear integral sequence by a pseudo-random digital process.

The fundamental design problem lies in the definition of an integral sequence and a digital selection system by which $N_a(t)$ and $N_b(t)$ meet the following requirements:

(i) Each must possess a zero-mean Gaussian amplitude probability density function with the same r.m.s. value.

(ii) Each must possess a power spectrum half-Gaussian in shape, and with the same r.m.s. width.

(iii) The two variables must be uncorrelated.

Having found a system which satisfies these conditions, a 'quantized' approximation to a Rayleigh fade may be produced by using each variable to amplitude modulate one of the quadrature signal sources and then adding the modulated signals.

3 Gaussian Modulator

The function of a Gaussian modulator is to generate one of the variables $N(t)$ and to amplitude modulate one of the quadrature signal sources with it.

3.1 Basic Principle

The theoretical circuit is shown in Fig. 1(a) and is based on an n-stage binary shift register which is fed with a random bit stream. Each register stage operates a changeover switch which connects either the input signal or its inverse, to a weighted summing resistor. The voltage at the output of the summer is given by:

$$V_0 = V \sin \omega t \sum_{i=1}^{n} a_i z_i \qquad (4)$$

where a_i is a weighting factor, z_i is selected randomly by the bit stream and may be $+1$ or -1 with equal probability. As each bit travels through the register it will contribute to the output a voltage which changes at each clock pulse as shown in Fig. 2, thus effectively amplitude modulating the input signal with a stepped waveform. The relative amplitude of the stepped waveform, at successive clock pulses is determined by the series $\pm(a_1, a_2, a_3, \ldots, a_i, \ldots, a_n)$, the sign depending on whether a 'one' or a 'zero' is travelling through the register. By suitable selection of the values of a_i this waveform can be made to approximate to any arbitrary curve.

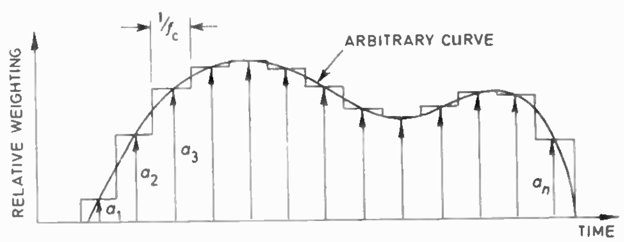


Fig. 2. 'Single bit' response of circuit of Fig. 1.

It is shown in the Appendix that the amplitude distribution of the output V_0 will approach Gaussian for any arbitrary weightings of the register stages, provided n is large. It is reasonable therefore to begin by considering the problem of achieving a Gaussian-shaped power spectrum. Assuming that the frequency spectrum of the random bit stream itself is essentially flat over the frequency range considered, (i.e. d.c. to several times the fading rate) then the frequency spectrum of the output V_0 will be determined solely by the transfer function generated by the switching process. By use of the convolution theorem it may be shown that the single pulse response of the system approximates to the impulse response provided that n is large, and in the Appendix it is shown that the criterion for a Gaussian-shaped power spectrum is met when the impulse response of the weighted register stages is of the form:

$$f(t) = \exp(-\tfrac{1}{2}(t/t_1)^2). \qquad (5)$$

Thus the required Gaussian amplitude probability density function and the Gaussian-shaped power spectrum may be obtained simultaneously by weighting the register stages according to the ordinates of a Gaussian curve (i.e. equation (5)). A certain degree of quantization error must be accepted (as with any analogue/digital conversion), and Fig. 3 indicates that a close approximation to a Gaussian-shaped pulse may be obtained by using a linear approximation, particularly if there are an even number of attenuator stages, which are therefore weighted:

$$\pm 1, \pm 2, \pm 3, \pm 4, \ldots, \pm \tfrac{1}{2}n, \pm \tfrac{1}{2}n, \ldots, \pm 4, \pm 3, \pm 2, \pm 1.$$

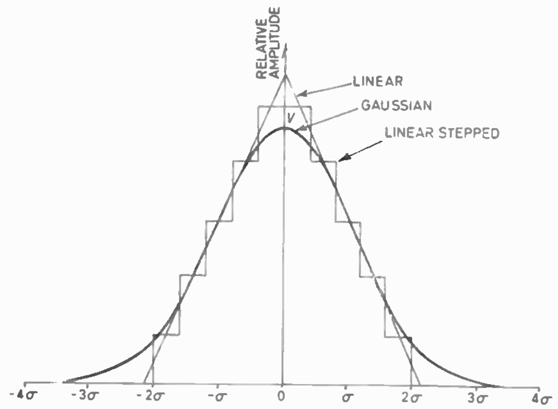


Fig. 3. Linear and stepped approximations to a Gaussian pulse.

The above method for generating the modulating functions is very convenient, as, with the addition of a modulo-2 feedback path, the shift register may be used as a chain code generator (c.c.g.) and so provide its own pseudo-random binary stream.[12] Uncorrelated bit streams for $N_a(t)$ and $N_b(t)$ may be generated by using different modulo-2 feedback configurations on the two chain code generators.

3.2 Optimum Number of Attenuator Stages

The theoretical limit on the accuracy of the simulator depends entirely on the number of stages used in the digitally programmed attenuators. The probability distribution and the auto-correlation function of the envelope of V_0 were computed for a 10-stage linearly weighted attenuator and Figs. 4(a) and (b) show the results. The approximations to the Gaussian curve are adequately accurate for most purposes, particularly when it is noted that the addition of two such vectors in quadrature will reduce non-systematic errors still further.

Assuming a linear weighting of 10-stage attenuators, the resulting Rayleigh distribution is composed of over 950 unique discrete vectors.

3.3 Practical Implementation

The practical implementation of Fig. 1(a) is shown in Fig. 1(b), where the changeover switches are replaced by on/off f.e.t. analogue switches, and then a phase-inverted signal added at a fixed level of half the peak variable signal.

The nominal values of the weighting resistor are chosen from a standard 5% series such that the weighting is optimized between a Gaussian curve and the inverse linear series:

$$1 : \tfrac{1}{2} : \tfrac{1}{3} : \tfrac{1}{4} : \tfrac{1}{5} : \tfrac{1}{5} : \tfrac{1}{4} : \tfrac{1}{3} : \tfrac{1}{2} : 1.$$

The number of stages in each chain code generator must obviously be equal to or greater than the number of attenuator stages, and in the latter case the attenuator stages must be connected to immediately successive stages of the c.c.g.

4 Rayleigh Fading Channel

The complete Rayleigh fading channel (see Fig. 5) is achieved by phase shifting the input baseband signal to produce two equal-amplitude versions of the input signal in phase quadrature, subjecting each to the Gaussian modulation process described above, and then linearly summing the two modulated quadrature signals.

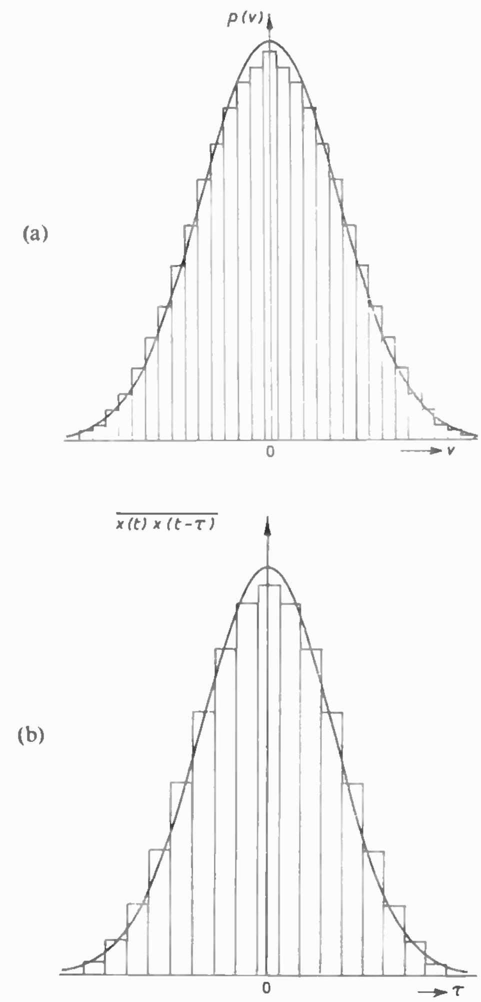


Fig. 4. Computed amplitude p.d.f. (a) and autocorrelation function (b) of a 10-stage linearly-weighted Gaussian modulator.

4.1 Phase Quadrature Network

To obtain the phase quadrature signals, it is convenient to use a passive network of the type described by Dome.[13] With 5% tolerance components there was no difficulty in obtaining a differential phase shift of 90° ± 3° and an amplitude characteristic that is flat within ±0·5 dB over the band of 250 Hz to 4 kHz for which the instrument was initially designed. Little difficulty is anticipated in extending this bandwidth if required.

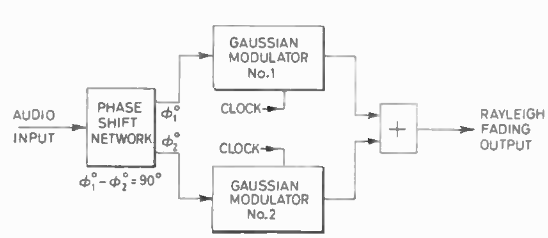


Fig. 5. Rayleigh channel—block diagram.

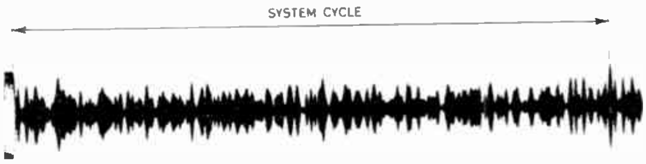
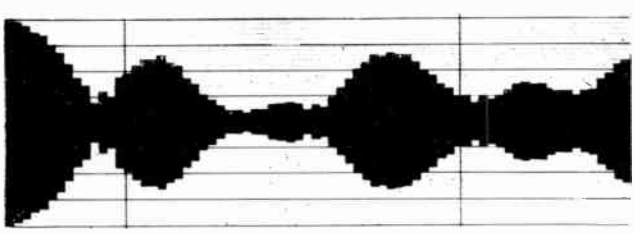


Fig. 6. Fading envelope of a single Rayleigh fading channel using two 10-stage chain code generators.

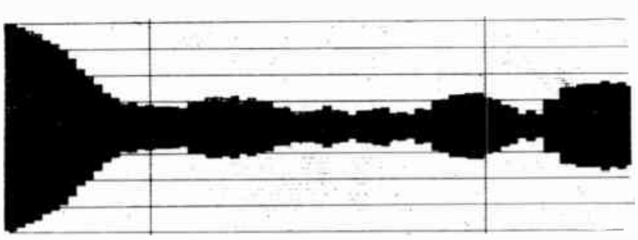
4.2 Chain Code Generator Considerations

The two c.c.g. circuits which control the two attenuators in a Rayleigh channel must generate uncorrelated pseudo-random bit streams and therefore two different feedback configurations must be used, each giving maximal-length codes. A complete system cycle of the Rayleigh fading output from such a circuit, using two 10-stage registers, is shown in Fig. 6. At a fading rate of 30 fades/second this cycle repeats after 2 seconds.

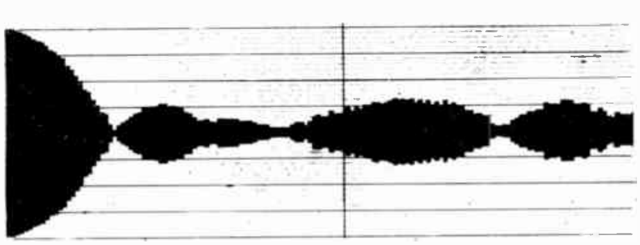
It is possible to increase the cycle length very considerably by using c.c.g.s of unequal length. The total cycle time of a single maximal length chain code generator of p stages is (2^p-1) clock pulses, and if two chain code generators of p and q stages are started together it may be shown that the overall cycle time (i.e. the time before the system returns to the initial condition) is $(2^p-1)\times(2^q-1)$ clock pulses provided 2^p-1 and 2^q-1 have no common factors. By using one 10-stage and one 13-stage c.c.g., the fading sequence in the prototype simulator lasts 5 hours at 30 fades/second.



(a) Rayleigh channel 1.



(b) Rayleigh channel 2.



(c) Combined channels.

Fig. 7. As Fig. 6 with expanded scale showing definition in time and amplitude domains.

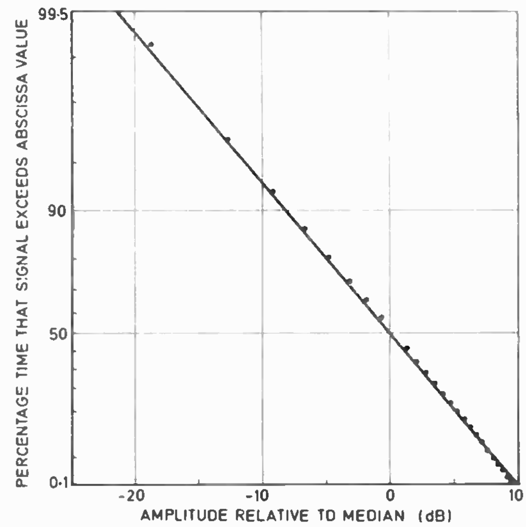


Fig. 8. Measured amplitude cumulative distribution function of combined channels.

The definition of the fading waveform in the time domain can be improved by clocking the two chain code generators in one channel on opposite phases of the clock, and since for most measurements two Rayleigh fading channels are combined (see equation (1)) a further improvement can be obtained by driving the four chain code generators in turn from a four-phase clock. With this arrangement there is an average of about 60 samples per fade cycle. The combination of two Rayleigh channels also results in an improvement of the amplitude quantization and the final output may be composed of over 3700 unique discrete vectors. This reduction in quantization error is clearly indicated in Fig. 7, which shows simultaneous u.v. recordings of two Rayleigh channels and their combined output.

4.3 Accuracy of Rayleigh Output

The cumulative distribution function of the fading envelope was determined by measuring the percentage of time for which various predetermined levels were exceeded during part of the fading sequence. The results are shown in Fig. 8; the plotted points representing the measured values and the straight line representing the Rayleigh distribution. The agreement between the measured points and the Rayleigh distribution is extremely good at all amplitude levels. Similarly Fig. 9 shows the measured power spectrum and the theoretical Gaussian-shaped power spectrum (continuous curve), and again the agreement is adequately accurate for simulation purposes.

A further check on the accuracy of the Rayleigh output may be obtained by measuring the relative number of crossings of the fading envelope through various levels. The results are shown in Fig. 10 along with the theoretical curve (which assumes a Rayleigh amplitude probability density function and a Gaussian-shaped power spectrum). Once again the agreement is very good at all amplitude levels.

Accurate measurement of the phase distribution presents practical difficulties and therefore a qualitative

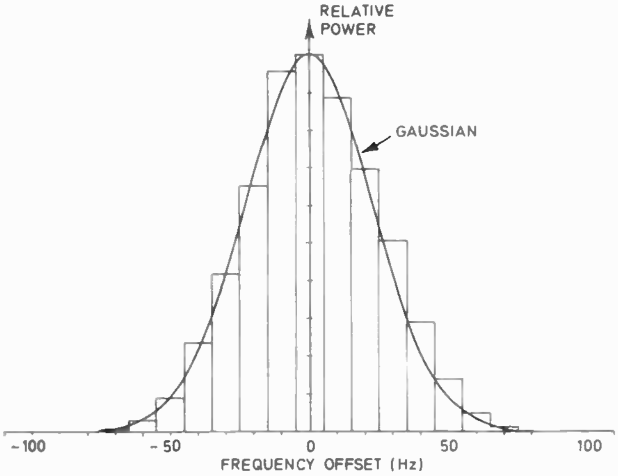


Fig. 9. Measured power spectrum of combined channels (33 fades/second).

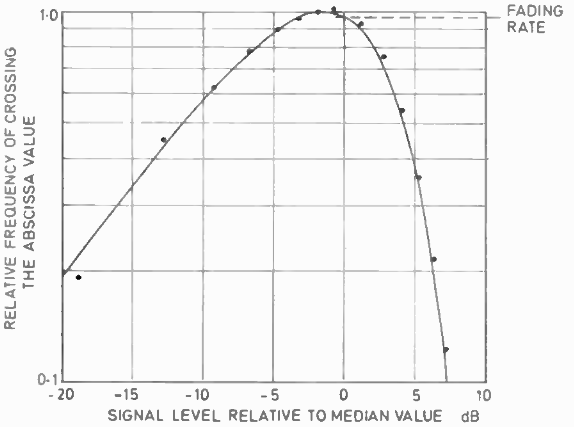


Fig. 10. Measured frequency of crossing of various levels (combined channels).

check was made with the aid of a polar plot. This may be generated by resolving the output signal into its sine and cosine components and then using the envelopes to control the X and Y deflections of an oscilloscope. The fading waveform moves in discrete amplitude steps and hence each level generates a spot of light, the transitions between levels being invisible. Thus a 'time' exposure photograph results in a convenient polar plot of both the phase and amplitude variations. The fine definition of the 'grid' pattern formed by the discrete levels and the extremely uniform phase distribution are clearly seen in the polar plot of Fig. 11(a). Figure 11(b) shows a small fraction of the fading cycle during which

(a) Large number of fade cycles. (b) Small number of fade cycles.

Fig. 11. Polar plots.

the deflection voltages were low-pass filtered to produce 'tails' showing the direction of movement.

The f.e.t. switching transients and the quantization noise produced by the discrete amplitude steps gives rise to unwanted spurious components separated from the baseband frequency by multiples of the c.c.g. clocking frequency. Measurements showed that all such spurious components are at least 30 dB below the wanted signal, and can usually be ignored. The discrete level changes of the quadrature signals generate instantaneous phase changes in the output, which may possibly produce anomalous results when the fading machine is used in conjunction with a phase-coherent signalling system (such as a d.p.s.k. modem).

The system of successive chain code generator clocking reduces this effect to a minimum and it was calculated that the mean phase change per master clock pulse is 7·6 degrees, which should be adequately accurate for most purposes but may limit the usefulness of the machine in some very specific applications. The use of more switched attenuator stages would of course reduce the effect.

4.4 Advantages of the Digital Quadrature Attenuation Method

The advantages of the digital quadrature attenuation method of generating a Rayleigh fading signal may be summarized as:

(i) The characteristics of the fading signal in both the amplitude and frequency domain may be made to approximate a true Rayleigh fade to any arbitrary degree of accuracy.

(ii) All signal processing takes place at baseband frequency and the limitations of the audio channel are largely those of simple m.o.s.f.e.t. switches and operational amplifiers used in inverting or summing modes. The system is therefore very stable and repeatable.

(iii) The r.m.s. and peak values of the signal level are accurately defined, enabling the maximum advantage to be taken of the dynamic range of analogue circuits.

(iv) The fading rate is accurately defined and dependent only on the clock rate. It may be varied easily, manually or by programming, over a wide range.

(v) The method offers facilities not achievable by existing methods; for instance a fading pattern may be identically repeated at a different fading rate, making it easy to identify and investigate the effects of specific fading phenomena (e.g. the rapid rate of change of phase in fading troughs).

(vi) The circuit can be constructed entirely from standard i.c. components, and is eminently suitable for low-cost quantity production.

5 Auxiliary Circuit Functions

5.1 Audio Delay

The path time delay T of equation (1) is usually simulated by the use of an audio delay line. Lumped-

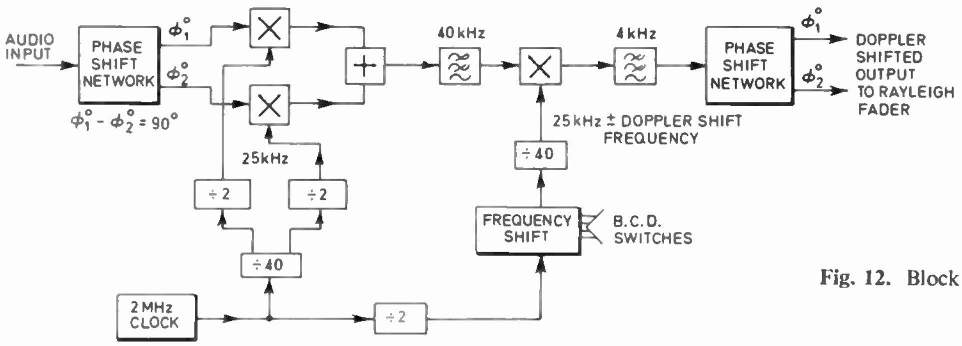


Fig. 12. Block schematic of Doppler shift circuit.

constant delay lines are not economical in this application but fortunately an m.o.s. charge-coupled device known as the 'bucket-brigade' delay line has recently become available.[14] Although the samples obtained by the authors lacked reliability, sufficient work was carried out to show the method is capable of providing an efficient solution to the problem. Two 32-stage chips were used and gave acceptable results up to a delay of about 8 ms, but for longer delays distortion of the higher audio input frequencies is not tolerable and the bandwidth must be limited. Longer delays or wider bandwidths, or both, may be obtained by using more stages. The delay time is controlled by a clock waveform, which in the fading simulator is obtained from a fixed-frequency multivibrator followed by a binary divider, allowing switched delays of 1–16 ms in binary steps.

5.2 Doppler Shift Circuit

A block schematic of the basic Doppler shift circuit is shown in Fig. 12. A phase-shifting network splits the input signal into two phase-quadrature versions which are subsequently modulated by two 25 kHz phase-quadrature square waves in m.o.s.f.e.t. chopper modulator circuits. The modulator outputs are then added and low-pass filtered to produce the lower sideband of a single sideband suppressed carrier signal which is then translated back to baseband by demodulation with a carrier which is offset from 25 kHz by the required Doppler shift. The offset carrier is generated from a 1 MHz source using a digital frequency shifting method developed by one of the authors (J.D.R.). The frequency shift may be varied from ±0·1 to 9·9 Hz, in 0·1 Hz steps, and ±1 to 99 Hz in 1 Hz steps. A disadvantage of the prototype circuit is that it causes amplitude modulation of the baseband signal (to a depth of about 10%) with the second harmonic of the Doppler shift frequency. This is probably acceptable in most cases but can be reduced by further development.

5.3 Noise Generator

A commercial semiconductor noise generator was used—probably an unwise choice as the power generated in a 4 kHz bandwidth is extremely low. The noise generated is amplified and low-pass filtered to a bandwidth of approximately 4 kHz.

6 Complete Equipment

6.1 Circuit Configuration

Figure 13 shows the block schematic of the complete equipment. The fading rate, Doppler shift and delay are determined (in absolute terms) by the frequency of clock oscillators and hence may be generated (or measured) to any degree of accuracy required. A Wien

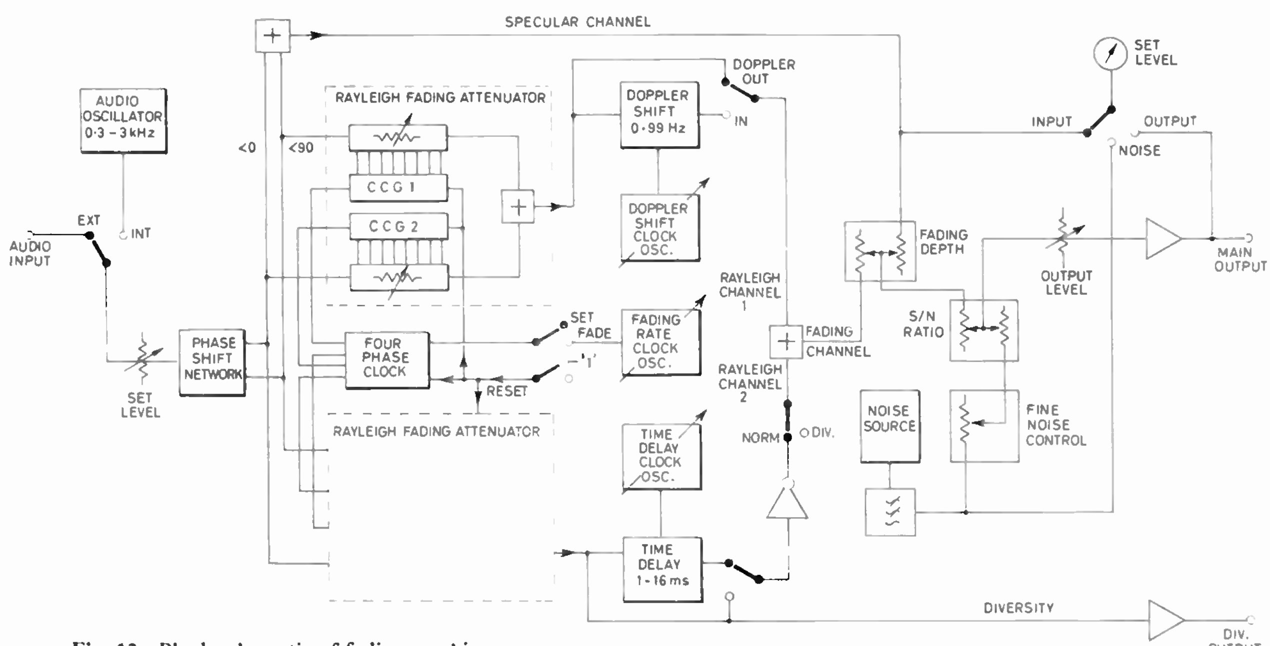


Fig. 13. Block schematic of fading machine.

bridge oscillator covering the range 300 Hz to 3 kHz provides a convenient signal source for demonstration or tone tests.

When the fading machine is in its 'Normal' mode the two fading channels are combined at the same r.m.s. level (i.e. $A_1 = A_2$ in equation (1)) and then added to the specular channel via a switched differential attenuator which allows selection of either the combined fading signal, the specular signal, or a weighted sum of the two in 10 dB steps from -30 dB to $+30$ dB giving a Riceian distribution. The differential attenuator is designed so that the total r.m.s. output remains constant for all attenuator settings (assuming that the two attenuator inputs are uncorrelated).

The Riceian signal is then added to the noise in a similar differential attenuator. The signal-to-noise ratio (measured in a 4 kHz equivalent noise bandwidth) may be varied in 10 dB steps from -30 dB to $+30$ dB. A fine control of 10 dB in 1 dB steps is provided by a separate attenuator in the noise source.

6.2 Operation of the Controls

The instrument is designed to be simple to operate, and is set up by first switching the meter switch to 'Input' and then adjusting the input level control until the meter pointer lies on the 'Set Level' mark. No further setting-up adjustments are necessary and all circuits are operating near to their optimum dynamic range for all control settings.

With the mode switch at 'Normal', the combination of both Rayleigh channels (plus all the auxiliary functions) is available at the main output socket. In the 'Diversity' position, only channel 1 is provided (with noise and specular components and Doppler shifted if required) at the main output, while channel 2 is available separately at the diversity output. Thus two independent uncorrelated Rayleigh fading sources are simultaneously available for the testing of dual diversity systems. The fading rate is continuously variable over the range 2 to 2000 fades per minute.

When the fading switch is in the 'Set' position all four c.c.g.s are preset to the all-one's condition and the four-phase clock to its starting condition, thus allowing an identical fading pattern to be repeated when the switch is returned to 'Fade'.

Intelligent use of the controls permits the simulation of many of the subsidiary effects that occur on h.f. and other links. For example, periodic fading[15] may be simulated by setting the fading switch to 'Set' and with a Doppler shift inserted the main output will consist of the sum of a Doppler shifted signal (channel 1) and an unshifted signal (channel 2). The fading rate of the resulting signal is equal to the Doppler frequency.

7 Practical Applications

The primary function of the machine was envisaged as enabling standard measurements of the performance of a communications system in noise[3,16,17] to be carried out easily and quickly by technicians as a routine part of a development programme and also to enable causes of inaccuracy in a communication channel to be investigated in detail. As an example, measurements have been carried out in the author's laboratory of the performance of a 32-tone m.f.s.k. telegraphy system (duplicating to some extent the work of Reference 16) and a 6-tone, two-element m.f.s.k. system. The measured results confirm those in the reference and agree with theoretical analysis to within about 1–1½ dB, except in the case of fast fading. The anomalous effects of fast fading were investigated in some detail, leading to a clearer concept of the causes of such errors (confirming the predominant effect of the fast phase changes occurring at low amplitude minima). A theoretical analysis was derived and confirmed experimentally. It is noteworthy that the same theoretical approach applied to a 6-oscillator fading simulator of the type described in Section 2.1, suggested that the incorrect power spectrum of that system could cause an apparent error in the fading rate (by this criterion) of nearly 2 to 1.

The prototype simulator has found some applications which were not originally envisaged. For instance it has been used to modulate an h.f. signal generator in order to compare the dynamic a.g.c. characteristics of two radio receivers. It has also proved extremely useful in demonstrating graphically and easily to non-technical people and to trainee operators and engineers the aural effects of various h.f. phenomena. The slow-time polar display illustrated in Fig. 11(b) is useful in the study by mathematics students of classic two-dimensional 'random walk' problems.

8 Conclusions

The authors consider that the aims of the project as outlined in Section 1 have been adequately fulfilled. The machine as constructed is portable and simple and convenient to use. It is also considerably more accurate and more flexible than any equipment of a similar nature known to the authors. For some specialized applications, such as the detailed investigation of some phase-modulation signalling systems, the occurrence of discrete phase changes in the output may prove a limitation on accuracy, but it is doubtful if this will prove a serious drawback in most cases.

9 Acknowledgments

The majority of the theoretical calculations, computer solutions, and the initial circuit development of the Rayleigh fading system, were originally carried out by one of the authors (F.M.E.S.) as a B.Sc. project at the University of Southampton, which was the subject of a detailed report.[18] The authors wish to acknowledge the permission given by the University to publish extracts from that report, and would like to thank Dr. J. A. Betts in particular for supervising the project, and the Department of Electronics as a whole for the facilities that were provided.

Subsequent work, supervised by Mr. R. D. Neale, in the Development Section of the Communications Engineering Department of the UK Foreign and Commonwealth Office, resulted in the construction of the complete laboratory prototype with the extended auxiliary facilities.

The authors also wish to thank the Director of Communications of the Foreign and Commonwealth Office and the Controller of HM Stationery Office for permission to publish the paper.

10 References

1. Law, H. B., Lee, F. J., Looser, R. C. and Levett, F. A. W., 'An improved fading machine', *Proc. Instn Elect. Engrs*, **104**, Part B, pp. 117–47, 1957.
2. Beck, D. and Betts, J. A., 'Fading machine for the simulation of the ionosphere', *Electronic Engng*, **37**, pp. 74–9, February 1965.
3. CCIR, 'H.F. Ionospheric Channel Simulators', XIIIth Plenary Assembly CCIR, Geneva 1974, Vol. III, Report No. 549.
4. CCIR, 'Fading of Radio Signals received via the Ionosphere', XIIIth Plenary Assembly CCIR, Geneva 1974. Vol. VI, Report No. 266–3.
5. Schwartz, M., Bennett, W. R. and Stein, S., 'Communication Systems and Techniques' (McGraw-Hill, New York, 1966).
6. Dasgupta, K. K. V., 'Statistical analysis of fading of a single downcoming wave', *J. Atmos. Terr. Phys.*, **18**, pp. 265–75, 1960.
7. Rice, S. O., 'Statistical properties of a sine wave plus random noise', *Bell Syst. Tech. J.*, **27**, pp. 109–57, January 1948.
8. Bradley, P. A., Eccles, D. and King, J. W., 'Some effects of the equatorial ionosphere on terrestrial h.f. communication', *Telecommunication J.*, **39**, No. 12, pp. 717–24, December 1972.
9. CCIR, 'Special Problems of H.F. Radio Communication Associated with the Equatorial Ionosphere', XIIIth Plenary Assembly CCIR, Geneva 1974, Vol. VI, Report No. 343–2.
10. CCIR, 'Multipath Propagation on H.F. Radio Circuits', XIIIth Plenary Assembly CCIR, Geneva 1974, Vol. III, Report No. 203.
11. Goddard, F. E., 'Description of a Long Element Two Tone Radio Telegraph System', Signals Order No. 6232 RDS Technical Minute No. SCL/87, RAF West Ruislip 1962/63.
12. Sterling, J. T., 'An Introduction to Pseudo Noise Codes and Correlators', General Electric Technical Information Series No. R62 DS D34, 5th May 1962.
13. Dome, R. B., 'Wideband phase shift networks', *Electronics*, **19**, pp. 112–5, December 1946.
14. Sangster, F. L. J., 'The "bucket-brigade delay line", a shift register for analogue signals', *Philips Tech. Rev.*, **31**, No. 4, pp. 99–110, 1970.
15. Davies, K., 'Ionospheric Radio Propagation', National Bureau of Standards Monograph 80, April 1965.
16. Wheeler, L. K. and Law, H. B., 'The Piccolo 32-tone Telegraph System Subject to Noise and Fading', GPO Research Report No. 20879 (March 1964).
17. Ridout, P. N. and Wheeler, L. K., 'Choice of multi-channel telegraph system for use on h.f. radio links', *Proc. Instn Elect. Engrs*, **110**, No. 8, pp. 1402–10, 1963.
18. Sladen, F. M. E., 'H.F. Ionospheric Channel Simulator', B.Sc. Project Report. Dept. of Electronics, University of Southampton, May 1973.

11 Appendix: Mathematical Analysis

11.1 Amplitude Probability Density Function of the Gaussian Modulator

Consider the function of Fig. 1(c). The output $N(t)$ consists of the arithmetic sum of the weighted digits x_i, in the register stages. Thus:

$$N(t) = \sum_{i=1}^{n} x_i$$

where $x_i = \pm a_i$ with equal probability.

All the x_i are independent random variables with expected value 0 and variance $(a_i)^2$, therefore from the central limit theorem the sum $x_1+x_2+\ldots+x_n$ has a distribution that approaches Gaussian with zero mean. Thus the required Gaussian amplitude distribution will be obtained for any arbitrary weighting of the register stages provided n is large.

11.2 Relation between Fading Rate and Clock Rate

Consider the 'single-pulse' response of the attenuator to be an ideal Gaussian-shaped pulse of unity amplitude. Thus:

$$f(t) = \exp\left(-\tfrac{1}{2}(t/t_1)^2\right) \tag{6}$$

the Fourier transform of which is given by

$$\mathrm{F}(\mathrm{j}\omega) = \sqrt{2\pi t_1}\exp\left(-\tfrac{1}{2}(f/f_1)^2\right) \tag{7}$$

where

$$f_1 = 1/2\pi t_1.$$

Now the energy spectral density $G(\omega)$ is given by $\mathrm{F}(\mathrm{j}\omega)^2$ hence

$$G(\omega) = 2\pi t_1^2 \exp - (f/f_1)^2. \tag{8}$$

Hence $G(\omega)$ is Gaussian with an 'r.m.s. frequency':

$$f_{\mathrm{r.m.s.}} = f_1/\sqrt{2}. \tag{9}$$

Observation of Fig. 3 shows that for a linear approximation to the Gaussian pulse the base of the pulse is approximately 4·4 standard deviations wide. For the function $N(t)$ in the time domain this corresponds to a width of $(n+1)/f_c$ where f_c is the clock frequency to the c.c.g. and n is the number of attenuator switches. The standard deviation of the Gaussian pulse is t_1 and therefore,

$$4{\cdot}4t_1 = (n+1)/f_c. \tag{10}$$

Substitution of equation (10) into equation (9) gives

$$f_{\mathrm{r.m.s.}} = \frac{4{\cdot}4f_c}{2\pi(n+1)\sqrt{2}} \simeq \frac{0{\cdot}5f_c}{(n+1)} \tag{11}$$

and since the fading rate (as usually defined) is given[4] by $f_r = 1{\cdot}47f_{\mathrm{r.m.s.}}$ then

$$f_r = 0{\cdot}728f_c. \tag{12}$$

Note that the master clock operates at $4f_c$.

11.3 Effect of Bit Stream Spectrum

The above calculations assume that the output spectrum is unaffected by the spectrum of the bit stream itself. In fact the Fourier transform of the output is given by

$$\mathrm{S}(\mathrm{j}\omega) = \mathrm{F}(\mathrm{j}\omega).\mathrm{H}(\mathrm{j}\omega) \tag{13}$$

where $\mathrm{H}(\mathrm{j}\omega)$ is the spectrum of the random bit stream in the register. This consists of a line spectrum with spacing $f_c(2^n-1)$, and a $(\sin x/x)$ shaped envelope with nulls at $f_c, 2f_c, 3f_c, \ldots$.

The first null is at $f_c = 2(n+1)f_{\mathrm{r.m.s.}}$, and so if $n \gg 1$ it can be assumed that $\mathrm{H}(\mathrm{j}\omega) = 1$ for all significant values of $\mathrm{F}(\mathrm{j}\omega)$.

Manuscript first received by the Institution on 3rd January 1975, in revised form on 11th December 1975 and in final form on 27th April 1976. (Paper No. 1744/Com. 136).

Letter

From: **G. May,** C.Eng., M.I.E.R.E.

A Pioneer of Airborne Radio

The section of the Institution's Golden Jubilee issue of *The Radio and Electronic Engineer* dealing with the history of the airborne radiotelephone[1] did not, I feel, give credit to the pioneering work of the Royal Aircraft Establishment teams concerned with airborne communications, particularly to the v.h.f. system development which was under the technical direction of Dr. A. C. Bartlett from 1935–39.

Since the RAE work has been described elsewhere[2,3] it is not the purpose of this letter to dwell solely on this point but rather, before the memories pass into obscurity, to recall also the considerable part played by Dr. Bartlett in the development of the theory of electronics and networks.

A. C. Bartlett's contributions were published during the period 1920–35 in journals such as the *Journal of Scientific Instruments*, *Experimental Wireless* and the *Philosophical Magazine*. It was however through his book 'The Theory of Electrical Artificial Lines and Filters,'[4] published in 1930, that he was to become best known and to which so many references have subsequently been made by those applying his bi-section theorem and his work on continuants. It is interesting to note that references[5,6] have appeared as recently as 1973 and in widely separated parts of the world, surely a tribute to the fundamental nature of his work. An example of how he was very much in advance of his time was the application of hyperbolic functions in conjunction with the transmission (*ABCD*) matrix to the solution of active circuit problems involving cascaded thermionic valve amplifiers[7] in 1930. Whilst visiting him in 1968 I mentioned how widely known his bi-section theorem had become, particularly in the USA, and was interested to hear that he regarded his paper on 'Intermodulation in audio amplifiers'[8] as being of greater importance. His final publication, 'On the geometry of rectangular waveguides'[9] appeared in 1948.

The part that he played and its importance in the field of airborne radio communications is perhaps best shown by the following extract from 'A Historical Summary of The Royal Aircraft Establishment 1918–1948'.[10]

'The decision to develop a complete system of air and ground VHF short range radio telephone communication equipment was taken at an early stage and proved to be an instance of outstanding foresight, particularly on the part of the late Wing Commander D. H. de Burgh, the Head of the Department during the development period. For some two years before 1938 a team including J. E. Clegg, R. Cockburn, D. W. Fry, L. Holt Smith, W. A. Johnson, D. G. Reid, J. Stewart and J. H. Mitchell working under Dr. A. C. Bartlett had carried out pioneer work in the hitherto unused frequency band of 100 to 124 Mc/s, but the threat of war led to extreme pressure of work, and in 1938 the experimental transmitters and receivers were quickly developed into forms suitable for production. The airborne transmitter-receiver, known as TR.1133, was in production in small numbers in September 1939 and fitting in Fighter Command had advanced sufficiently during 1940 for full operational use to be made of it during the Battle of Britain, with the old MF equipment available in reserve. An equal contribution to the success of the system was made by the ground based equipment, including the 50 watt transmitter T.1131 and its associated receiver R. 1132, direction finding equipment, aerial masts and heads and forward relay stations with remote control over land lines. Even the earliest aircraft transmitter-receiver was of advanced design, being a true radio-telephone with quartz crystal frequency control and requiring no attention by the pilot other than selection of one of four operating channels by push-button remote control. The design was quickly followed up by a replacement set the TR.1143, the prototype again being made in the Department. The new set was a better production proposition and contained valves more suited to the purpose, which the industry had now had time to develop. An early model of the TR.1143 was taken to the United States of America in 1940, and by agreement the first American VHF Fighter set was based on the TR.1143 and made mechanically and electrically interchangeable. A significant difference was the extension of frequency coverage to 156 Mc/s with the American expansion needs in view. Many elements of the short-range voice communication system described survived the war period and continued in use, the remainder having been superseded by improved units. The use of VHF equipment for this service rapidly spread to all Commands of the Royal Air Force, the Fleet Air Arm and Air Forces of most European countries, and later became the international standard practice for Civil Aviation'.

I only became acquainted with Dr. Bartlett in the last few years of his life and I am indebted to his nephew Mr. A. P. J. Goff for his assistance in providing much of the limited biographical details we have been able to collect.

Permission to quote from the official history of RAE is also gratefully acknowledged.

G. May

Royal Aircraft Establishment
Radio and Navigation Department
Farnborough, Hants GU14 6TD

2nd August 1976

Biographical Notes on A. C. Bartlett

Born in Cambridge on 3rd August 1893 and graduated in 1919 from Emmanuel College, Cambridge. Served in the Royal Corps of Signals in World War I. Worked for MO Valve Company from 1920 to 1935. In 1935 he was awarded the degree of Doctor of Science by Cambridge University in recognition of this research work and in the same year joined the Experimental Wireless Department of RAE. About 1940 transferred to TRE Swanage (later absorbed into RRE Malvern). Died at Fivehead, Taunton, Somerset 27th July 1974.

References

1. Brinkley, J. R. 'Fifty years of mobile radio', *The Radio and Electronic Engineer*, **45**, No. 10, pp. 557–58, October 1975.
2. Gates, B. G., 'Aeronautical communications', *J. Instn Elect. Engrs*, **94**, Part IIIA, No. 11, pp. 78–80, 1947.
3. Gates, B. G., 'The heavens are filled with shouting', Chap. 9 of 'Laboratory of the Air' (HMSO, London, 1948).
4. Bartlett, A. C., 'The Theory of Electrical Artificial Lines and Filters' (Chapman and Hall, London, 1930).
5. Stevenson, J. K., 'Synthesis of wide band cascaded double-pole crystal lattice filters', *The Radio and Electronic Engineer*, **43**, No. 10, pp. 631–7, October 1973.
6. Szentirmai, G., 'Synthesis of multiple-feedback active filters', *Bell Syst. Tech. J.*, **52**, No. 4, p. 527, April 1973.
7. Bartlett, A. C., 'Multi-stage valve amplifier', *Phil. Mag.*, **10**, pp. 734–8, October 1930.
8. Bartlett, A. C., 'Intermodulation in audio-frequency amplifiers', *Wireless Engineer*, **12**, pp. 70–4, February 1935.
9. Bartlett, A. C., 'Geometry of rectangular waveguides', *Wireless Engineer*, **25**, pp. 202–10, July 1948.
10. Caunter, C. F., 'A Historical Summary of the Royal Aircraft Establishment 1918–1948', RAE Report Aero 2150A, 1949.

UDC 621.372.54:621.374

Iterative technique for designing non-recursive digital filter non-linear phase characteristics

Professor A. G. J. HOLT,
Ph.D., D.Sc., C.Eng., F.I.E.R.E.*

J. ATTIKIOUZEL, B.Sc., Ph.D.,†

and

R. BENNETT, B.Sc., M.Sc.*

SUMMARY

Iterative techniques based upon Remez's method can be used to design non-linear digital filters having small attenuation ripples in the stop band without appreciable distortion arising in the pass band.

* *Department of Electrical and Electronic Engineering, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, NE1 7RU.*

† *Formerly at the University of Newcastle-upon-Tyne; now at the Department of Electrical and Electronic Engineering, Portsmouth Polytechnic, Portsmouth PO1 3DJ.*

1 Introduction

Non-recursive digital filters are frequently associated with linear phase responses. However, in applications where a non-linear phase response is necessary the non-recursive filter structure may still be used. A design procedure based on optimization has recently been presented[1] for the design of such non-linear phase filters. In this paper we shall show that an iterative method based on the Remez exchange algorithm[2,3] can be used to overcome some major obstacles occurring using optimization.

2 Theory

The brief treatment of the theory given below is very similar to the theory section of reference 1. The frequency response of a non-recursive digital filter can be written as

$$H(\omega) = \sum_{K=0}^{N-1} d_k \exp(-jk\omega T) \tag{1}$$

where d_k are the delay line weighting coefficients which influence the shape of the frequency response and T is the sampling period. Initially consider the case when N is even. It is possible to control the phase response by considering the weighting coefficients as being the sum of two symmetric sets, one possessing even symmetry and the other odd, i.e.

$$d_k = a_k + b_k$$

and

$$d_{N-1-k} = a_k - b_k$$

where $k = 0, 1, 2, \ldots, (\frac{1}{2}N-1)$. Substituting for the new coefficients into equation (1), the frequency response becomes:

$$\begin{aligned} H(\omega) = \sum_{k=0}^{\frac{1}{2}N-1} \{a_k \exp(-jk\omega T) + \\ + a_k \exp(-j\omega T[N-1-k])\} + \\ + \sum_{k=0}^{\frac{1}{2}N-1} \{b_k \exp(-jk\omega T) - \\ - b_k \exp(-j\omega T[N-1-k])\}. \end{aligned} \tag{2}$$

By using trigonometric identities and expanding, equation (2) becomes

$$\begin{aligned} H(\omega) = 2\exp(-j\omega T[N-1]/2) \times \\ \times \left\{ \sum_{k=0}^{\frac{1}{2}N-1} a_k \cos(\omega T[k-(N-1)/2]) - \right. \\ \left. -j \sum_{k=0}^{\frac{1}{2}N-1} b_k \sin(\omega T[k-(N-1)/2]) \right\}. \end{aligned} \tag{3}$$

From equation (3) it can clearly be seen that the magnitude response is obtained by taking the absolute value of the real and imaginary parts.

The overall phase response may be considered as a linear function with a superimposed deviation. The linear function is dependent solely upon the filter order, while the deviation also depends upon the coefficients. It is convenient to define the phase response as the phase deviation from linear.

Similarly for N odd the transfer function may be written as

$$H(\omega) = \exp(-j\omega T[N-1]/2) \times \\ \times\left\{a_{\frac{1}{2}N-2} + 2\left[\sum_{k=0}^{\frac{1}{2}(N-3)} a_k \cos(\omega T[k-(N-1)/2] - \\ -j\sum_{k=0}^{\frac{1}{2}(N-3)} b_k \sin(\omega T[k-(N-1)/2]\right]\right\}. \quad (4)$$

To design a filter with a constrained phase response it is necessary to approximate the magnitude and phase deviation obtained from equations (3) and (4) to the ideal magnitude and phase responses, $D_i(\omega)$ and $\phi_i(\omega)$ respectively. Expressing these in terms of real and quadrature components we have

$$R(\omega) = D_i(\omega)\cos[\phi_i(\omega)] \quad (5)$$

and

$$Q(\omega) = D_i(\omega)\sin[\phi_i(\omega)] \quad (6)$$

respectively. The 'rectangular' shape is often chosen as the ideal low-pass magnitude response, while the ideal phase response is dependent on the particular application.

3 Approximation Problem

The approach suggested by Cuthbert[1] relies on approximating the absolute value of the real and imaginary parts of equation (3) to the ideal real and quadrature components using optimization. This approximation may also be done by the utilization of an iterative process based upon Remez's exchange algorithm.[2] It should be noted that by adopting either method no precise control is made over the phase error, however, by using the Remez technique the magnitude is guaranteed to be within a prescribed tolerance, whereas optimization does not give such a guarantee. In practice, it has been found that provided the magnitude tolerance is small the phase approximation is very good in the pass band but deviates from the ideal in the cut-off region. Provided this stop-band attenuation is large it is not necessary to approximate the phase in the stop-band region. In fact, no phase control is obtained since $D_i(\omega) = 0$ in this interval.

The magnitude response can be obtained from

$$|H(\omega)|^2 = H_R^2(\omega) + H_Q^2(\omega)$$

where $H_R(\omega)$ and $H_Q(\omega)$ are the real and imaginary approximations respectively. Hence, for an overall tolerance of $\lambda(\omega)$ it is necessary to realize each component to a tolerance of $\lambda(\omega)/\sqrt{2}$.

4 Approximation Method

The technique first requires the formulation of an error term for each component. This can be expressed as

$$E_R(\omega) = [R(\omega) - H_R(\omega)]W_R(\omega) \quad (7)$$

for the real component, and

$$E_Q(\omega) = [Q(\omega) - H_Q(\omega)]W_Q(\omega) \quad (8)$$

for the quadrature component

where $W(\omega)$ is the weighting function.

If $E_R(\omega)$ and $E_Q(\omega)$ are of equiripple nature possessing N turning points in the region $0 < \omega \leqslant \omega_s/2$, then the best approximation is obtained. It must be stressed here that if N is odd the real error function will possess an additional turning point. If the error function, $E(\omega)$, is constrained such that it is equal to unity at its extreme points and by letting the weighting function be

$$W(\omega) = \frac{\sqrt{2}}{R_p}$$

in the pass-band region and

$$W(\omega) = \frac{\sqrt{2}}{R_s}$$

in the stop-band region, then the overall magnitude response will be within the required tolerances R_p and R_s in the pass and stop-band respectively. A closer phase approximation is obtained by letting

$$W_R(\omega) = \frac{\sqrt{2}}{[R_p \operatorname{abs}(R(\omega))]}$$

and

$$W_Q(\omega) = \left[\frac{\sqrt{2}}{R_p \operatorname{abs}(Q(\omega))}\right]$$

for the weighting function of the real and quadrature components. For this case the percentage tolerance of each component is constant.

The approximation may be achieved by the application of the Remez exchange algorithm to the error term $E(\omega)$. Remez's algorithm is realized by an iterative process that at the kth stage:

1. Fits $E(\omega) = (-1)^j$ at some set of points $(\omega_j)_k$ to obtain an intermediate approximation $E_k(\omega)$.
2. Finds the locations $(\omega_j)_{k+1}$ of the extrema of the error curve for $E_K(\omega)$.
3. Goes to step (1), substituting $(\omega_j)_{k+1}$ for $(\omega_j)_k$.

Difficulties may arise in applying this technique due to precision and ill-conditioning. The ill-conditioning of the parameter fitting equations may be alleviated by using orthogonal polynomials for the basis of equation (4). For the case considered here, $H(\omega)$ is an orthogonal-type function and thus it is not necessary to transform it to a Chebyshev-type polynomial. Considering the process in more detail, the first step is to choose a set of extrema (ω_j) for the starting points ($k = 1$). The zeros of the Chebyshev polynomials $T_N(\omega)$ may be used, for example, although this is an arbitrary choice since convergence is guaranteed for any starting point.

It is now possible to equate the error function

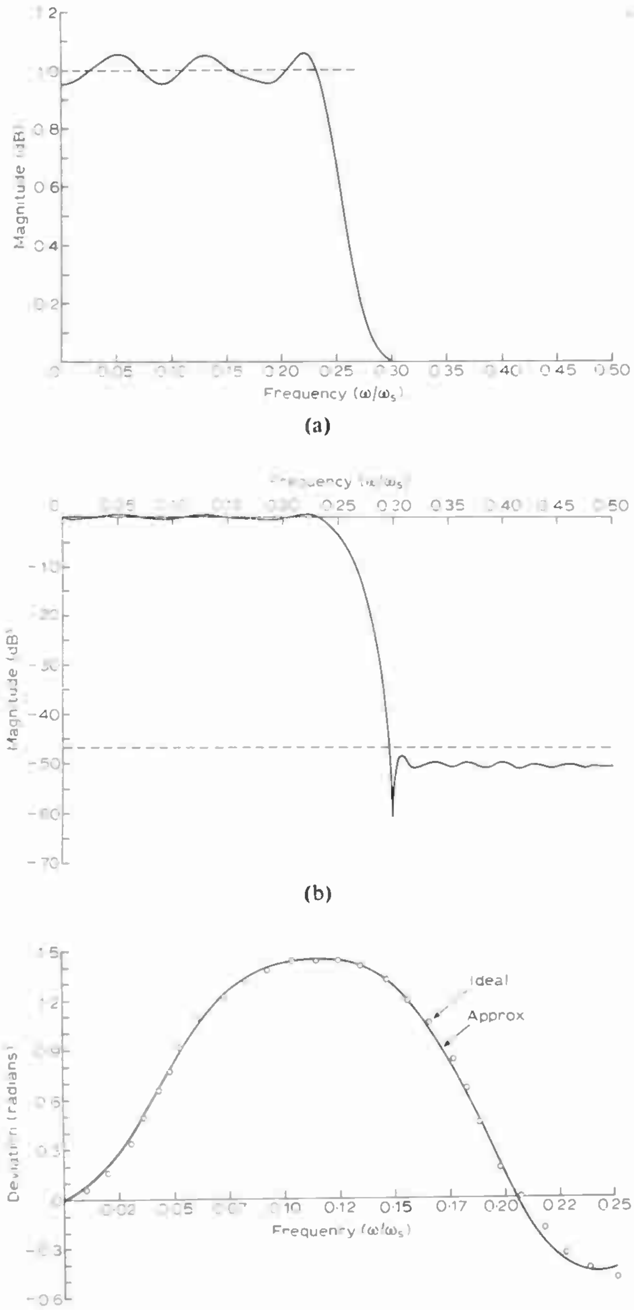
$$E(\omega) = (-1)^j = W(\omega)[Q(\omega) - H(\omega)] \quad (9)$$

at the N guessed points. The set of N equations are now solved for the unknown coefficients using a method based upon Gaussian elimination.

The next step is the determination of the extrema points of the new function.

5 Locating the Extrema

Due to the prior knowledge that the function $E(\omega)$ has N extrema in the range $0 < \omega \leqslant \omega_s/2$, it is more efficient to determine the turning points using a scan technique rather than root solving the derivative. The root solving method becomes impractical for high-order filters.



(a)

(b)

(c)

Fig. 1. Design example of 32nd-order delay line.
(a) and (b) Frequency response showing magnitude approximation.
(c) Phase deviation approximation and ideal.

The searching technique locates the extrema by first testing for a change of sign in the derivative, for linear increments of the frequency variable. The precise location of the extrema may then be obtained by a simple iteration round these points. Due to the properties of the orthogonal polynomials the extrema are nearly linearly spaced, which allows the function to be sampled only $5N$ times with a very low probability of missing a turning point. If all the extrema points are not found then the process may be repeated using a higher scan rate. Finally, it is necessary to test the process after each iteration to see if the function lies within a prescribed error. An error at an extrema of 10^{-5} to 10^{-6} is usually acceptable.

It must be pointed out here that the overall magnitude response $|H(\omega)|$ will not necessarily be of an equiripple nature. However, the response will lie within the prescribed tolerance at all points.

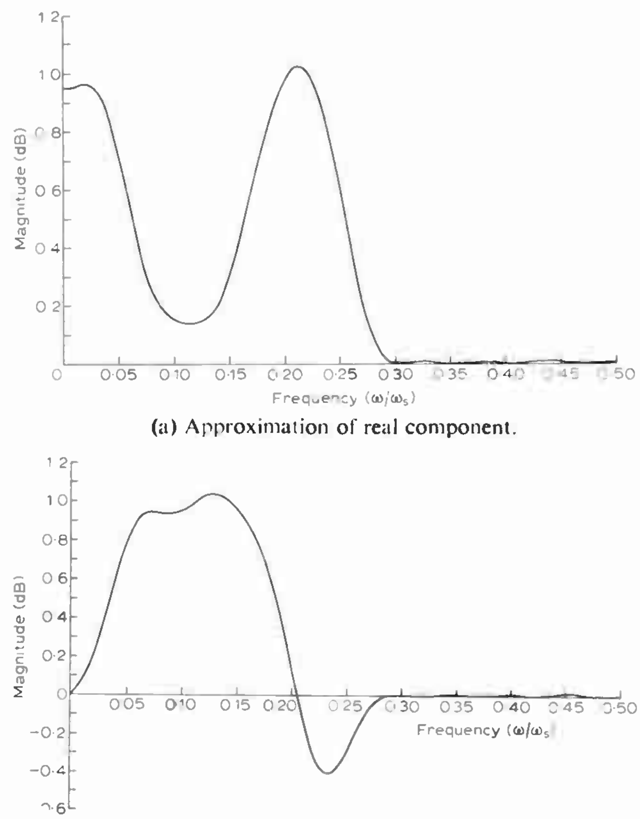
6 Design Examples

Consider the following filter specifications:

pass-band ripple = 0·08
stop-band ripple = 0·005

The ideal phase deviation was obtained by an approximation to a set of data points[1] kindly supplied by Dr. L. G. Cuthbert.

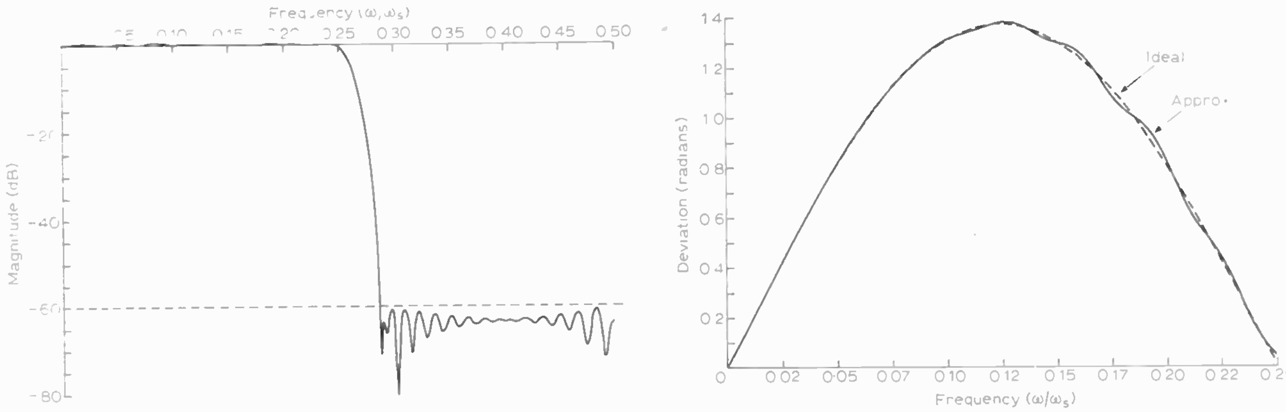
A 32nd-order delay line was successfully used to approximate the required filter. The amplitude response and phase deviation from the ideal response are plotted in Fig. 1 (a), (b) and (c) respectively. As can be seen from the Figures, a 32nd-order filter satisfies both the pass and stop-band tolerances, and in addition, a good approximation to the phase deviation is obtained throughout the pass-band region. The real and quadrature approximations are shown in Fig. 2 (a) and (b). These approximations illustrate the effect of the phase response changing by almost 90° in the pass-band interval. Furthermore, the quadrature approximation crosses the x-axis during the pass-band interval, this necessitates the first stop-band ripple to be positive. Using this technique it is possible to design high order filters. This is illustrated by considering the following filter specifications.



(a) Approximation of real component.

(b) Approximation of quadrature component.

Fig. 2.



(a) Frequency response magnitude approximation.

(b) Phase deviation approximation and ideal.

Fig. 3. Design example of 64th-order delay line.

pass-band ripple = 0·05
stop-band ripple = 0·001
phase deviation = $1{\cdot}37 \sin (2\omega T)$

The amplitude response for a 64th-order filter with the above specifications is shown in Fig. 3(a). The phase deviation of this filter is given in Fig. 3(b).

7 Conclusions

The versatility of this design approach allows a considerable range of different filters to be designed for given pass and stop-band tolerance specification and phase response in contrast to optimization routine. This is achievable because a trade-off between order and cut-off frequencies is possible for given ripple magnitudes. In addition, by use of the weighting factors, different responses may easily be obtained. Using the Remez approach, convergence is guaranteed and is virtually independent of any initial guess. Although it is necessary to express the phase response in an analytic form, it is possible to approximate a set of measured points by the use of standard curve fitting routines.

Remez's exchange algorithm is very efficient, requiring approximately only 3 second of c.p.u. time on an IBM 370/168 for a 24th-order filter. In contrast to optimization routines small ripples are obtainable in the stop-band without any distortion occurring in the pass-band. The procedure has been successfully tested for responses with 100 dB rejection in the stop-band and with phase deviations of 90°.

8 Acknowledgments

The research project presented in this paper was financed by the Royal Signals and Radar Establishment, Ministry of Defence.

Thanks are due to Mr. R. Kell for interesting and stimulating discussions.

9 References

1. Cuthbert, L. G., 'Optimizing non-recursive digital filters to non-linear phase characteristics', *The Radio and Electronic Engineer*, 44, No. 12, pp. 645–51, December 1974.
2. Acton, F. S., 'Numerical Methods That Work' (Harper and Row, London, 1970).
3. McClellan, J. H. and Parks, T. W., 'A unified approach to the design of optimum f.i.r. linear phase digital filters', *Trans. IEEE on Circuit Theory*, CT-20, No. 6, pp. 697–701, December 1973

Manuscript first received by the Institution on 11th August 1975 and in revised form on 28th November 1975. (Short Contribution No. 180/CC265.

UDC 681.322.64

Integrator design for a differential analyser

W. FORSYTHE, M.Sc.*

and

S. L. HOUSEMAN, B.Sc.†

Based on a lecture to the East Midland Section of the Institution in Leicester on 17th October 1974.

SUMMARY

The paper sets out to examine various integration algorithms, suitable for implementation in the form of serial binary arithmetic, and concludes that only first-order integration is practicable at the present time and that only first-order prediction is necessary or desirable.

An examination is made of the errors inherent in digitizing the main functions of the traditional analogue computer (integration, multiplication, division) and some conclusions reached about the feasibility of building a digital differential analyser given the current state of technology.

**Department of Electronic and Electrical Engineering, University of Technology, Loughborough, Leicestershire, LE11 3TU.*

†Formerly at the University of Loughborough; now with IBM UK Laboratories, Hursley Park, near Winchester.

List of symbols and abbreviations

a_k, a_y, etc.	representational error in K, Y, etc.
A	maximum amplitude of a sinewave
d.l.a.	double length accumulator
E_{Ii}	ith order integrator error
E_{Imi}	maximum value of $\lvert E_{Ii}\rvert$
E_{pi}	ith order predictor error
E_{pmi}	maximum value of $\lvert E_{pi}\rvert$
f_i	frequency of the input signal
f_s	sampling frequency
F_{du}	fractional error in U with a double length accumulator
F_{su}	fractional error in U with a single length accumulator
G	integrator gain
h	sampling interval Δt
$k = Gh$	
K	computer representation of k
N	number of samples per cycle
Q	the value of one least significant bit
$r = 2\pi f_i/f_s$	
R	minimum acceptable resolution
s.l.a.	single length accumulator
U	the output of the integrator multiplier, Δz
$y(t)$	input to the integrator at time t seconds
$y(t+h)$	input to the integrator at time $t+h$ seconds
y', y''	first, second derivative of $y(t)$
Y	computer representation of $y(t)$, or the integral of $y(t)$
$z(t)$	integrator output

1 Introduction

There has always been interest in the problem of replacing the conventional analogue computer by some digital equivalent, and this has heightened in recent years with the advent of transistor-transistor logic (t.t.l.) and other technical advances.

The designer of a digital analogue computer, or digital differential analyser (d.d.a.) to give it its classical name, has a choice to make at the beginning. He must decide whether to base his design upon incremental transmission or total transmission. The computer must of course operate in the sampled-data mode in either case, transmitting information between computing units at specific sampling instants, but the information may be at the latest value of the complete variable x, or only the change Δx since the previous transmission. These are referred to respectively as total transmission and incremental transmission (the latter may in fact involve not the transmission of differences but differences of differences[1]).

The favoured approach until recently has almost invariably been incremental transmission[2] because it involves only a small number of bits, say one, two, or three. This, however, imposes a rate limit on the system which is more severe than may at first be suspected. Consider for simplicity an incremental system transmitting only one bit, the least significant, of value Q. The sampling rate is N samples/second and we wish to handle a sine wave $x = A \sin \omega t$. The maximum rate of change of x is $A\omega$ and the maximum rate of change that the system can handle is NQ. The sampling rate necessary for the frequency ω is therefore

$$N = A\omega/Q$$

Let us stipulate a minimum resolution R such that

$$A = RQ$$

and we may now write

$$N = R\omega$$

If $\omega = 6 \times 10^3$ (approximately 1 kHz) and $R = 10^4$ we are not demanding anything extraordinary from the system but N is sixty million samples a second! Since all calculations must be carried out within $1/N$ s, the basic clockrate must be considerably in excess of 60 MHz —a very high figure for an unremarkable demand. We can of course bring this value down by transmitting more bits but we then sacrifice the simplicity of our arithmetic circuits, which was the main reason for adopting this approach in the first place. One may cling to the incremental approach by applying greater subtlety[1] but the original simplicity is invariably lost.

It is worthwhile, therefore, to investigate the other possibility, total transmission, hitherto considered impossibly slow though for no good reason.

The remainder of this paper is concerned exclusively with total transmission and examines the design and performance of suitable digital integrators, upon which the performance of the computer as a whole is crucially dependent. In addition, the errors involved in multiplication and division are examined, bearing in mind that both devices use predicted quantities.

2 Serial Digital Integrator Design

Total transfer could in theory be implemented in either parallel or serial fashion but it is not technically feasible at the moment to transfer, say, 24 bits in parallel between computing units through a patchpanel or

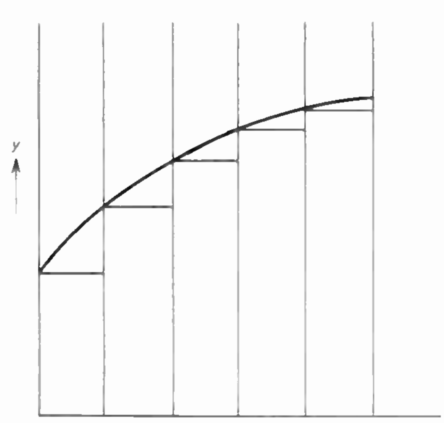


Fig. 1. Integration using rectangle summation

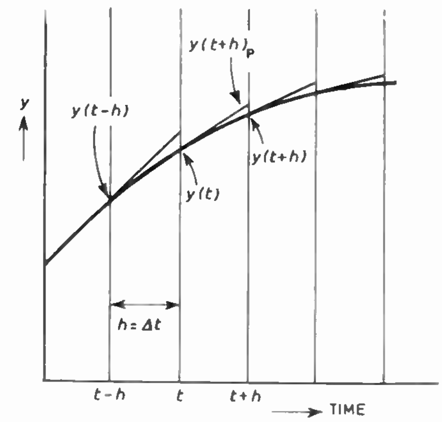


Fig. 2. Integration using trapezoids

switching system, so it must be done in series. This has often led people to condemn the approach out of hand as too slow because it must require a transmission time and therefore a sampling interval appreciably longer than that required for incremental transmission. However, although the sampling interval must be greater, the performance is not *necessarily* inferior on that account.

The speed and accuracy of computation is determined principally by the integrators so integrator design is the central issue.

It must be borne in mind that, while the general-purpose digital computer possesses only one processor, the d.d.a. has many. For economic reasons therefore each one must be unsophisticated and cheap, and so the integrator designs considered here are all of a simple nature.

2.1 The simplest digital integrator could use rectangle summation (Fig. 1) to approximate the integral. This, the zero order or 0-integrator, would require only the most rudimentary circuitry and would therefore be very attractive economically, but its performance leaves a lot to be desired so we pass on here to a somewhat higher degree of complexity and consider first the implementation of trapezoidal integration (Fig. 2).

This involves us immediately with the problem of prediction. The integrator, in common with most other units in the computer, takes a substantial part of one sampling period to calculate the result for that step, so that although the input data are relevant to time t the output is not available until time $(t+h)$ where h is the length of the sampling interval. In a circuit such as that shown in Fig. 3 the output of the summer in these circumstances would be a mixture of values from different sampling instants. To prevent this we must use in the calculation not the value of y at time t but the predicted value of y at time $(t+h)$, denoted $y(t+h)_p$.

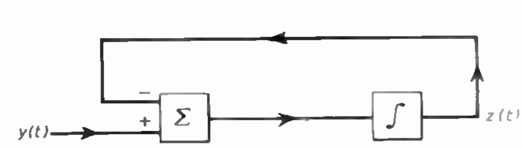


Fig. 3. A computing circuit to generate a simple lag

The basis of the prediction must be simple (again for economic reasons) which suggests a first-order predictor. This means storing the value $y(t-h)$ and, with the current value $y(t)$, predicting on a straight line the value $y(t+h)_p$. This integrator, then, uses first-order prediction with first-order integration and is referred to as a 1.1 integrator. Since

$$y(t+h)_p - y(t) = y(t) - y(t-h)$$

i.e. we are predicting the same change over the next interval as over the last, then

$$y(t+h)_p = 2y(t) - y(t-h) \qquad (1)$$

Using serial arithmetic this algorithm is particularly simple to realize (see Appendix 1) and $y(t+h)_p$ is available in nanoseconds, from the arrival of $y(t)$, should it be explicitly required for the rest of the computation.

The area of a trapezoid (Fig. 2) is given by the expression

$$\Delta z(t+h) = h\{y(t+h)_p - \tfrac{1}{2}(y(t+h)_p - y(t))\} = h\{y_s(t)\}$$

and therefore, substituting from equation (1) above,

$$y_s(t) = \tfrac{1}{2}(3y(t) - y(t-h)) \qquad (2)$$

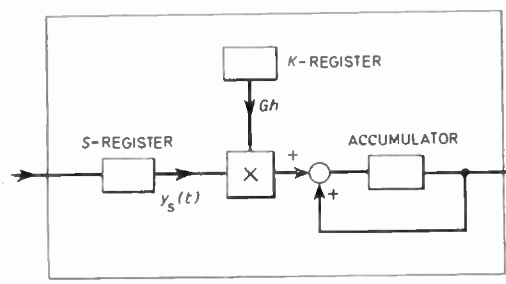


Fig. 4. Data flow within the integrator

The integrator determines $y_s(t)$ and multiplies this by $k = Gh$, where G is the gain of the integrator, to obtain the increment Δz. This is accumulated to give the output z:

$$z = \sum \Delta z = \sum k y_s = \sum G y_s h \simeq G \int y \, dt$$

A schematic diagram of the data-flow in the integrator is shown in Fig 4.

2.2 The trapezoidal integrator employs first-order prediction with first-order integration. To improve on this one might try second-order prediction with second-order integration, referred to as a 2.2 integrator. The orders of prediction and integration do not have to be the same but if we are storing a number of values of the incoming variable we might as well use them all for both purposes, or so it could be argued.

The second-order predictor requires the algorithm

$$y(t+h)_p = 3y(t) - 3y(t-h) + y(t-2h)$$

while second-order integration (on its own) requires

$$\Delta z(t) = \frac{h}{12}\{5y(t) + 8y(t-h) - y(t-2h)\}$$

Combining these to produce the algorithm for the 2.2 integrator that we want, gives us

$$\Delta z(t+h) = h\{y_s(t)\}$$

where

$$y_s(t) = \tfrac{1}{12}(23y(t) - 16y(t-h) + 5y(t-2h)) \qquad (3)$$

The numbers involved in this algorithm are extremely unattractive (in contrast to those of the previous algorithm) and in fact any integration routine above the order of one produces these unpleasant fractions. They are both time-consuming and expensive to implement in serial arithmetic.

2.3 Unlike integration routines, prediction routines of any order present no special problem. We could therefore attempt to improve on the 1.1 integrator by using second-order prediction but maintaining first-order integration to avoid the awkward fractions, i.e. a 2.1 integrator. This results, as before, in

$$\Delta z(t+h) = h\{y(t+h)_p - \tfrac{1}{2}(y(t+h)_p - y(t))\} = h\{y_s(t)\}$$

but now

$$y(t+h)_p = 3y(t) - 3y(t-h) + y(t-2h)$$

This leads to

$$y_s(t) = 2y(t) - \tfrac{3}{2}y(t-h) + \tfrac{1}{2}y(t-2h) \qquad (4)$$

The coefficients in this expression are clearly very convenient.

Appendix 3 illustrates a method of deriving prediction and integration algorithms of any order.

2.4 The notion of retroactive correction on the following step is an attractive one which can easily be applied and may be thought of as correcting on the next step the predictive error on any order of prediction.

Consider, for instance, an integrator using zero-order prediction with retroactive correction (a 0.R integrator). The first value of Δz is Δz_1, given by (see Fig. 5)

$$\Delta z_1 = A_1 C_1 B_0 A_0$$

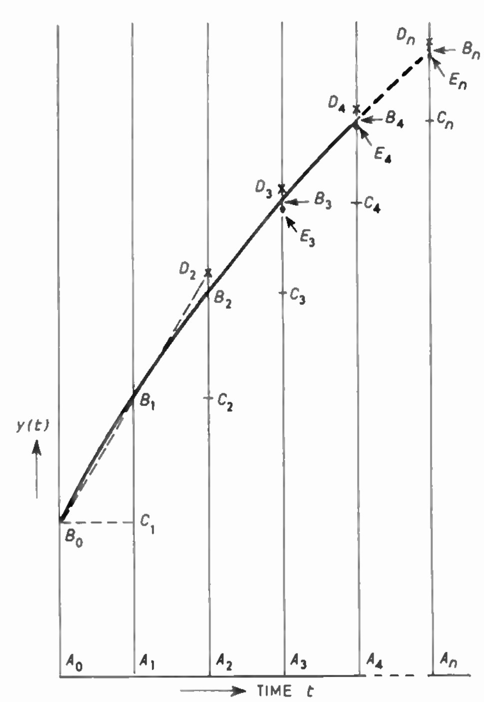


Fig. 5. Retroactive correction routines

(This must always be the value of Δz_1 whatever the order of prediction.) The second value Δz_2 is given by

$$\Delta z_2 = A_2 C_2 B_1 A_1 + C_1 B_1 B_0$$

where the second term is the correction. But

$$C_1 B_1 B_0 = C_2 D_2 B_1$$

where D_2 is $y(2h)_p$ based upon a first-order prediction from B_0 and B_1.

The algorithm required for the 0.R integrator is therefore the same as that for the 1.1 integrator and it can now be seen that the 1.1 integrator, as described in Section 2.1, does not accumulate error due to prediction, as one might expect, except on the last (the nth) step. The true curve $y(t)$ is represented as a series of n straight line segments jointing the points $B_0, B_1, B_2, \ldots, B_{n-1}, C_n$, the last point being the only one that does not lie on $y(t)$. The zero-order predictor when used with retroactive correction therefore produces error only on the last step so that most of the error in the computation is due to the use of the first-order integration routine.

2.5 An integrator using first-order prediction with first-order integration and retroactive correction (a 1.1.R integrator) will on the first step calculate (referring to Fig. 5 again)

$$\Delta z_1 = A_1 C_1 B_0 A_0$$

on the second step

$$\Delta z_2 = A_2 D_2 B_1 A_1$$

and on the third step

$$\begin{aligned}\Delta z_3 &= A_3 E_3 B_2 A_2\\ &= A_3 D_3 B_2 A_2 - E_3 D_3 B_2\\ &= A_3 D_3 B_2 A_2 - B_2 D_2 B_1\end{aligned}$$

So Δz_2 has now been tailored to the curve, being given the area $A_2 B_2 B_1 A_1$. Likewise Δz_3 will retroactively assume the area $A_3 B_3 B_2 A_2$ due to the action of Δz_4. Note, however, that this leaves Δz_1 uncorrected and still given by the area $A_1 C_1 B_0 A_0$. By way of compensation the last point on the curve is now D_n (not C_n) so that Δz_n is more accurate than is the case with the 0.R integrator. Using the 1.1.R integrator, then, the curve $y(t)$ is represented as the series of straight line segments joining the points $B_0, C_1, B_1, B_2, B_3, \ldots, B_{n-1}, D_n$. The accuracy gained on the nth step, the area $C_n D_n B_{n-1}$ may or may not balance the inaccuracy on the first step, the area $C_1 B_1 B_0$, depending upon the shape of $y(t)$. What is clear, however, is that the extra costs involved in undertaking the second-order prediction are not worthwhile, though this statement is intended to apply only to prediction on the integrand and not necessarily to prediction on other quantities around the computer.

The algorithm required to implement the 1.1.R integrator may be derived as follows:

The area calculated by the first-order integrator on the ith step is

$$\begin{aligned}\Delta z_i &= A_i D_i B_{i-1} A_{i-1}\\ &= \frac{h}{2}(3y(t) - y(t-h))\end{aligned}$$

The correction term is the area

$$B_{i-1} D_{i-1} B_{i-2} = \frac{h}{2}(-y(t) + 2y(t-h) - y(t-2h))$$

and subtracting these we get the full algorithm for the 1.1.R integrator, namely

$$\Delta z(t+h) = h\{y_s(t)\}$$

where

$$y_s(t) = 2y(t) - \tfrac{3}{2}y(t-h) + \tfrac{1}{2}y(t-2h) \qquad (5)$$

Comparison of equations (4) and (5) shows that the 2.1 and the 1.1.R integrators, like the 1.1 and the 0.R integrators, are the same thing. Retroactive correction therefore may be achieved in general by increasing the order of prediction by one, but as has been shown above there is little point in going above the first-order predictor because what is gained on the swings is largely lost on the roundabouts. A further argument against the use of higher orders of prediction is the practical one that if the nth order predictor is to behave as presumed in the argument above then a means must be found of ensuring that in the first $(n-1)$ steps of the computation, when nth order prediction is not possible, the predictor behaves like all the $(n-1)$ lower-order predictors in turn (zero order on the first step, first order on the second, second order on the third, etc.). If this is not done the performance will undoubtedly be inferior to a simple first-order routine.

The perhaps surprising conclusion is therefore that a first-order predictor is best, all things considered, and that most of the error in computing Δz is due to the first-order integration. Higher orders of prediction bring the predicted points appreciably closer to the true curve, as can be seen on Fig. 5, but nevertheless do nothing to increase the accuracy of the integration.

2.6 The choice of the integration algorithm is not the only factor affecting the performance of the integrator. Having computed $y_s(t)$ this is multiplied by k (Fig. 4) and added to the accumulator, and one might debate whether the accumulator should be one word long or two, remembering that the full multiplier output is double length. Arguing at a superficial level, it is obviously 'safe' to accumulate the full multiplier output of $2n$ bits, where the word length is n bits, though transmitting to the following computing unit only the most significant n bits of the accumulator. However, is it necessary to have a double-length accumulator? Since the inaccuracy of the calculation, due to prediction, etc., will be a good deal greater than the value of the nth bit, might it not be the case that the lower half of a double-length accumulator is almost random and therefore not worth the cost of collecting?

A critical look at the situation shows that it is not only safe but very desirable to employ the double-length accumulator for several reasons. These are explained in Section 3.5.

The multiplier design is equally important. Any multiplier that is not faulty will of course give the same product as any other, but time is of the essence and the faster the multiplier works the better. Unfortunately the faster it works the more it costs, so we find ourselves on the horns of the usual dilemma, in this case a particularly painful one because so many multipliers are used in a d.d.a. (a small machine might contain forty or fifty).

2.7 One of the great advantages of the digital integrator over its analogue counterpart is its ability to vary the gain dynamically. The K-register (see Fig. 4) holds the quantity $k = Gh$ where G is the gain and h is the step length, a constant. There is no reason why this register should not be accessible from the patch-panel, as well as from the control console, and when this is done the integrator has not only the data or Y input but a gain or G input which enormously increases its versatility. A further benefit that flows from this arrangement is that one may now integrate with respect to any variable x, which may or may not be time. It is merely necessary to place in the K-register the quantity Δx, the difference between successive values of x (this can be arranged without trouble) and $z = \int y \, dx$ can be computed.

The initial condition of the integrator is normally keyed in from the console but can likewise be arranged so that the appropriate register may, at the programmer's discretion, be controlled from the patch panel.

2.8 One of the most thorny problems inherent in d.d.a. design is presented by discontinuous functions but it is essential to cope with them.

There are only two ways in which discontinuities may be generated within the computer, by the resetting of an integrator or by the operation of a switch, both of them brought about by a logical signal produced for the purpose. When either operation takes place, prediction on the following step must be suspended (one step for a first-order predictor, two steps for a second-order predictor, etc.) so that for one period prediction is of zero order in those computing units connected to the switch or integrator.

This is achieved in the MBD-24 (see Sect. 4) by generating a short pulse at a specific time in the sample interval when switching or resetting has taken place. The pulse is generated by the switch or integrator and detected, if it is present, by the following computing unit or units which then suspend prediction for one period. The system works very well in practice though, of course, it adds complexity and therefore cost to the design of the units.

The only alternative would seem to be one that requires the user to patch the switching or resetting signal to a 'prediction' input point on all following units and so directly suspend prediction when required. This might have some advantage from the point of view of design simplicity but would involve the operator in a lot of extra patching and mean a larger patch panel and is probably inferior for these reasons.

3 Sources of Error

The errors dealt with here are the per-step errors; that is, those produced in a computing unit at each step of the computation, as distinct from the total solution error. The relationship between these two is difficult to discuss in general terms because it must depend heavily upon the specific computing configuration.

The per-step computing errors may be classified under five headings:

1. Representation
2. Round-off
3. Prediction
4. Integration
5. Underflow

and these are dealt with in the following five sections. In addition, Section 6 deals with the effect of predictor error on multipliers and dividers.

3.1 Representation error is simply that incurred in using a given number of bits to represent the value of a variable. It must be present in any digital machine and can only be reduced by increasing the word length.

The magnitude of the representation error is equal to half the least significant bit, on average.

3.2 Round-off error is produced when we round off the output of a computing unit, such as a multiplier or a divider, to the length of one word. It is usual to round off towards zero, rather than just up or down to avoid the accumulation of positive or negative errors.

3.3 The predictor is the most obvious source of error and since prediction is required in almost all computing units, we look at zero, first- and second-order predictors.

Zero-order prediction is given by

$$y(t+h)_p = y(t) \tag{6}$$

and the error, defined as predicted value minus actual value, is

$$E_{p0} = y(t+h)_p - y(t+h) \tag{7}$$

Substituting from equation (6) into (7) and using a Taylor series expansion for $y(t+h)$

$$E_{p0} = -hy'(t) - \frac{1}{2!}h^2y''(t) - \frac{1}{3!}h^3y'''(t) - \dots \tag{8}$$

First-order prediction is given by

$$y(t+h)_p = 2y(t) - y(t-h)$$

Using Taylor series expansions for $y(t-h)$ and $y(t+h)$ we get

$$E_{p1} = h^2y''(t) - \frac{2}{4!}h^4y''''(t) - \dots \tag{9}$$

Second-order prediction is given by

$$y(t+h)_p = 3y(t) - 3y(t-h) + y(t-2h)$$

so that the error

$$\begin{aligned} E_{p2} &= y(t+h)_p - y(t) \\ &= -h^3y''' + \tfrac{1}{2}h^4y''''(t) + \dots \end{aligned} \tag{10}$$

To gain some feel for the meaning of equations (8), (9) and (10) let us suppose that the input to the predictor is a sinewave of frequency f_i. Since h is necessarily a small quantity, probably measured in microseconds, the terms in higher powers of h will be extremely small so that we may approximate the error expressions, without much loss over a wide band of frequencies, to

$$E_{p0} = -hy'(t) \tag{11}$$

$$E_{p1} = h^2y''(t) \tag{12}$$

$$E_{p2} = -h^3y'''(t) \tag{13}$$

The errors as given by equations (11) to (13) are obviously sinusoidal, like the inputs. Figure 6 shows the maximum amplitude of E_{p0}, E_{p1} and E_{p2} for a range of input frequencies f_i expressed in terms of the sampling frequency f_s. Considering, for instance, zero-order prediction we have that

$$|E_{p0}|_{max} = h|y'(t)|_{max} = E_{pm0}$$

$$y(t) = \sin(2\pi f_i t)$$

Therefore

$$|y'(t)|_{max} = 2\pi f_i$$

and

$$E_{pm0} = 2\pi h f_i$$

Since

$$h = 1/f_s$$

where f_s is the sampling frequency, we may write

$$E_{pm0} = 2\pi(f_i/f_s) = r \tag{14}$$

where r is directly proportional to the ratio of signal frequency to sampling frequency. For convenience $\log_{10} E_{pm0}$ is plotted against $\log_{10} r$ in Fig. 6.

Perhaps the most useful piece of information to be gleaned from inspection of Fig. 6 is the sort of improvement to be expected by increasing the order of the predictor.

From equations (12), (13) and (14)

$$E_{pm1} = r^2 \tag{15}$$

$$E_{pm2} = r^3 \tag{16}$$

where E_{pm} indicates the maximum value of $|E_p|$. Hence

$$\log_{10} E_{pm1} - \log_{10} E_{pm0} = \log_{10} r = \log_{10} E_{pm2} - \log_{10} E_{pm1}$$

That is to say, at any given value of r or signal frequency the change in error due to an increase of one order in the prediction is a decrease by a constant factor r (see Fig. 6). If we put the sampling frequency at something realistic, say 10^5 samples/second, and choose a second-order predictor as a compromise between cost and performance we have, from Fig. 6 or equation (16), that

$$f_i = \frac{10^5}{2\pi}(E_{pm2})^{1/3} \tag{17}$$

Hence if E_{pm2} is say $1/10^3$, then f_i may be as high as 1592 Hz, i.e. 63 samples/cycle;

if E_{pm2} is say $1/10^6$, then f_i may be as high as 159·2 Hz, i.e. 630 samples/cycle.

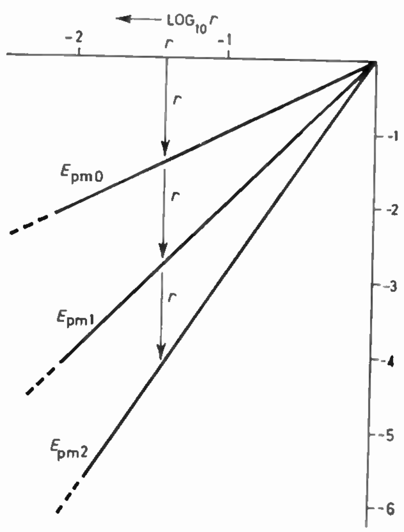


Fig. 6. Maximum per-step predictor error, E_{pm}, versus r for a sinewave input

3.4 Integration errors may be treated in a manner similar to that used for prediction errors.

The output of a true integrator operating over the sample period of h seconds is given by

$$z(t) = G\int_{t-h}^{t} y(t)\,dt + z(t-h) = G\{Y(t) - Y(t-h)\} + z(t-h) \tag{18}$$

where $Y(t)$ is the integral of $y(t)$.

Expanding $Y(t-h)$ in a Taylor series:

$$z(t) = G\{Y(t) - (Y(t) - hY'(t) + \frac{h^2}{2!}Y''(t) - \ldots)\} + z(t-h)$$

But

$$Y'(t) = y(t), \qquad Y''(t) = y'(t)$$

etc., so that

$$z(t) = G\left\{hy(t) - \frac{h^2}{2!}y'(t) + \frac{h^3}{3!}y''(t) - \ldots\right\} + z(t-h)$$
$$= K\left\{y(t) - \frac{h}{2!}y'(t) + \frac{h^2}{3!}y''(t) - \ldots\right\} + z(t-h) \tag{19}$$

where $K = Gh$.

The output of a first-order integrator is given by:

$$z(t) = z(t-h) + K\tfrac{1}{2}\{y(t) + y(t-h)\} \tag{20}$$

$$= z(t-h) + K\frac{1}{2}\left\{y(t) + y(t) - hy'(t) + \frac{h^2}{2!}y''(t) - \ldots\right\}$$
$$= z(t-h) + K\left\{y(t) - \frac{h}{2}y'(t) + \frac{h^2}{2\times 2!}y''(t) - \ldots\right\} \tag{21}$$

Subtracting equation (21) from equation (19) gives the per-step error due to the first-order integrator:

$$E_{I1} = K\left\{h^2y''(t)\left(\frac{1}{3!} - \frac{1}{2\times 2!}\right) + \ldots\right\} = -\frac{K}{12}h^2y''(t) - \ldots \tag{22}$$

The output of a second-order integrator is given by:

$$z(t) = z(t-h) + K\tfrac{1}{12}\{5y(t) + 9y(t-h) - y(t-2h)\} \tag{23}$$

and it may be shown in a similar manner that the error due to second-order integration is:

$$E_{I2} = -\frac{K}{24}h^3y'''(t) + \ldots \tag{24}$$

Note that equations (20) and (23) assume that the points $y(t)$, $y(t-h)$, etc., lie on the curve being integrated so that the errors E_{11} and E_{12} are due solely to the integration routines.

If the curve $y(t)$ contains no components which are of a frequency comparable to $1/h$, i.e. the number of samples per cycle is reasonably high, then the first terms in the series of equations (22) and (24) give good approximations to the errors.

To make these results meaningful in everyday terms let us suppose that the input to the integrator is $y = \cos \omega t$ and that the integrator gain is ω, so the output is $z = \sin \omega t$. To a first order of approximation the maximum per-step integration error, E_{Im}, for first- and second-order integrators is given by

$$E_{\mathrm{Im1}} = \frac{r^3}{12} \quad \text{and} \quad E_{\mathrm{Im2}} = \frac{r^4}{24}$$

(remember that $K = Gh = \omega h = \omega/f_s$).

A graph of $\log_{10} E_{\mathrm{Im}}$ against $\log_{10} r$ for these two cases is shown in Fig. 7, together with the curve for the zero-order integrator (for which $E_{\mathrm{Im0}} = r^2/2$) just for comparison.

If again we choose a sampling frequency of 10^5 samples/second and a first-order integration routine, then

$$f_i = \left(\frac{10^5}{2\pi}\right)(12E_{\mathrm{Im1}})^{1/3}$$

and if $E_{\mathrm{Im1}} = 10^{-3}$, then $f_i = 3643$ Hz
or 27 samples/cycle

if $E_{\mathrm{Im1}} = 10^{-6}$, then $f_i = 364$ Hz
or 270 samples/cycle.

With second-order integration then

$$f_i = \left(\frac{10^5}{2\pi}\right)(24E_{\mathrm{Im2}})^{1/4}$$

and if $E_{\mathrm{Im2}} = 10^{-3}$, then $f_i = 6267$ Hz
or 16 samples/cycle

if $E_{\mathrm{Im2}} = 10^{-6}$, then $f_i = 1114$ Hz
or 90 samples/cycle.

3.5 Underflow error occurs only in integrators that have a single-length accumulator. Let us suppose that the system is working with an n-bit word the least significant bit of which has the value Q $(= 2^{-(n-1)}$ because the most significant bit is 2^0). If the output of the multiplier within the integrator is of value less than Q then the single-length accumulator will record that as zero. (Remember that this may happen when the multiplier inputs are still quite significant.) Over a period of time therefore an error accumulates that may become appreciable.

Consider first the single-length accumulator (s.l.a.) scheme. Let true values be represented by small letters and machine representations by capitals.

The integrator multiplies together two numbers Y and K where

$$Y = y + a_y, \qquad K = k + a_k$$

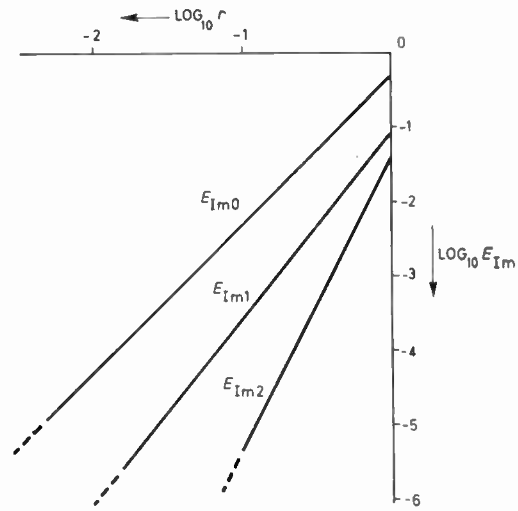


Fig. 7. Maximum per-step integrator error, E_{Im}, versus r for a sinewave input

a_y and a_k are the representation errors mentioned in Section 3.1. The output of the multiplier is $U (\equiv \Delta Z)$ and

$$U = Y * K = y * k + a_{su} = u + a_{su}$$

where $a_{su} \le 1$ l.s.b.

$$U = u(1 + F_{su})$$

where F_{su} is the fractional error in U and is given by

$$F_{su} = a_{su}/yk$$

Figure 8 shows contours of F_{su} on the y–k plane.

For the double-length accumulator (d.l.a.) scheme we have that

$$U = Y * K = y * k + y * a_k + k * a_y + a_y * a_k$$
$$= y * k + a_{du} = u + a_{du}$$

where

$$a_{du} = y * a_k + k * a_y + a_y * a_k$$
$$U = u(1 + F_{du})$$
$$F_{du} = \frac{a_k}{k} + \frac{a_y}{y}$$

(the product term $a_y * a_k$ being negligibly small). Given that on average $a_y = a_k = 2^{-24}$ then

$$\frac{1}{k} + \frac{1}{y} = 2^{24} * F_{du}$$

where

$$1 > \begin{Bmatrix} F \\ y \\ k \end{Bmatrix} > 2^{-24}$$

Bearing the following points in mind, it may be seen that the graph of $\log_2 k$ versus $\log_2 y$ takes the shape shown in Fig. 9:

For a given value of the parameter F,

(i) when $y = 1$ then

$$k = \frac{1}{(2^{24} * F) - 1} \simeq \frac{1}{2^{24} * F} \quad \text{for } F > 2^{-21} \text{ (say)}$$

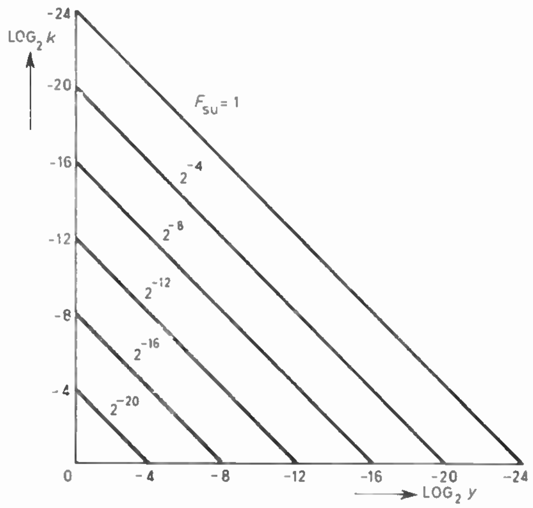


Fig. 8. Fractional error F_{su} as a function of k and y for a single-length accumulator

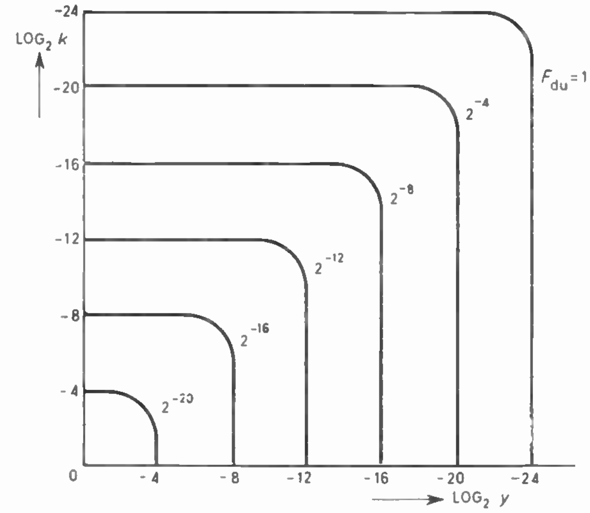


Fig. 9. Fractional error F_{du} as a function of k and y for a double-length accumulator

i.e. when $\log_2 y = 0$, then $\log_2 k = -24 - \log_2 F$.

(ii) when $y = k$, then

$$k = \frac{2^{-23}}{F} = y$$

i.e. $\log_2 y = \log_2 k = -23 - \log_2 F$.

(iii) the function is symmetrical with respect to y and k.

The contrast between Figs. 8 and 9 is striking and illustrates the improvement to be gained in using the d.l.a. scheme instead of the s.l.a.

In a typical computer situation the value of k is a constant, though it does not have to be (see Sect. 2.6), and the value of y, after reaching a peak declines towards zero.

It is as y gets small and the system approaches a steady state that the improvement with d.l.a. becomes important. Not only is the multiplier output U (or ΔZ) of much higher accuracy but the integrator no longer 'freezes'. As mentioned above, if the value of U falls below one least significant bit of the single-length word then on the s.l.a. system nothing is added in to the accumulator so the output freezes and will not change unless the input y increases and exceeds some minimum level. This freezing can have disastrous consequences on a computing circuit. Consider, for instance, the configuration of Fig. 10. A step change in x from say 0 to A will set both integrators off. Under d.l.a. w will eventually equal x and y will become zero, thus bringing integrator 02 to a halt at $z = f(A)$. Under s.l.a., however, integrator 01 may freeze as y becomes small so that y gets no smaller. If integrator 02 has not already frozen it will now simply run away until it overloads, and the circuit does not work. To avoid this with s.l.a. it is necessary to scale the circuit with run-away in mind and that in the end involves a loss of accuracy. It is this consideration that provides the most important argument against the use of s.l.a.

3.6 Prediction was discussed above, in Section 3.3, but what is the effect of predictor error on the output of a multiplier or a divider?

The derivation of the results is shown in Appendix 2 but they may be summarized as follows:

For a multiplier with first-order prediction on the inputs the error is

$$E = h^2(x''y + y''x) \tag{25}$$

If the prediction is performed on the output instead, then

$$E = h^2(x''y + y''x + 2x'y') \tag{26}$$

so that input prediction is the more desirable.

Second-order prediction on the inputs gives

$$E = h^3(x'''y + y'''x) \tag{27}$$

For the divider with first-order prediction on the inputs

$$E = \frac{h^2}{y^2}(x''y - y''x) \tag{28}$$

or predicting on the output

$$E = \frac{h^2}{y^2}\left(x''y - y''x - \frac{2y'}{y}(x'y - y'x)\right) \tag{29}$$

Comparison of equations (28) and (29) suggests that input prediction is again superior.

4 Discussion

The performance of the d.d.a. is inevitably measured against that of the analogue computer using the twin yardsticks of accuracy and dynamic range or bandwidth.

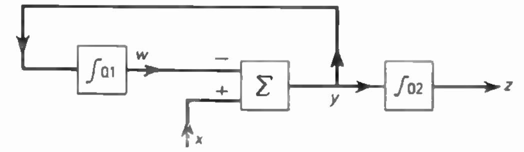


Fig. 10. A computing circuit that derives advantage from d.l.a.

Quantitative comparison would only be possible on specific computing circuits and one is then faced with the question of which circuits. Certain configurations would favour the differential analyser and others the analogue computer so be wary of results that could be 'arranged'.

By way of general or qualitative assessment, one can say that the accuracy of both types of computer falls off at high and at low frequencies, though for different reasons, of course. The low speed performance of the digital machine is not affected by drift but the values of Δz being computed are very small so that round-off error is relatively large, even using the double-length accumulator, though d.l.a., instead of s.l.a., brings about a big improvement, as the curves of Figs. 8 and 9 indicate. A greater word length would improve overall accuracy for a given sampling speed, particularly at the low frequency end.

The higher frequency errors could be reduced in three ways, by increasing the sampling rate, the order of the prediction routines (though not on the integrand, as explained in Section 2.4), and the order of the integration routines. The first two measures are quite feasible, in the present state of the art, but not the third.

To give our generalizations a point of contact with reality, the MBD-24, manufactured by Membrain Ltd., is a 24-bit machine using s.l.a. and a sampling rate of 10^4 samples/second (not 10^5 as used in the illustrative calculations of Section 3). The integrator gain range that is normally useable would be roughly 1 to 2000 and the accuracy of the output probably an order of magnitude better than we expect from the analogue computer. If therefore the design of the computer were updated to incorporate d.l.a., a sampling rate of 10^5, second-order prediction on multipliers, etc., and perhaps a 32-bit word, the performance would be impressive. As it stands, it is very useful and quite comparable with any analogue computer.

The complexity of the modern analogue computer is very much greater than would be the case with the proposed d.d.a., which need not be any more complex than the existing MBD-24. This means that both capital costs and running costs should be less for the differential analyser than for a comparable analogue computer (if one exists) or, to put it another way, the computing power for a given price should be much higher.

5 Conclusions

Many d.d.a.s have been built in the past (we have seen a list of over forty[4]) all, as far as we know, on the incremental principle but none has proved viable commercially, probably for technical reasons (highly accurate but very slow).

Technology has moved on, however, and it would now be possible to build a d.d.a. along the lines indicated in this paper with an excellent technical specification and very competitively priced. This is not to say that the company producing it would sweep the market. Indeed, the market in this country is so small and the major users in it so heavily committed to conventional analogue computers that the d.d.a., superior or not, could have a lean time of it for some years. Probably the only hope of the manufacturers recouping their development costs in a reasonable time is a successful invasion of either Europe or the US and this means considerable financial outlay before sales commence. It seems to the authors improbable therefore that any d.d.a., other than the existing one, is likely to find its way onto the market for some years to come despite the fact that an excellent machine could be built now.

6 References

1. Bywater, R. E. H., 'The second-order-difference integrator', *Proc. Instn Elect. Engrs*, **119**, No. 2, pp. 138–142, Feb. 1972.
2. Sizer, T. H. R., 'The Digital Differential Analyser' (Chapman and Hall, London, 1968).
3. Forsythe, W., 'Critique of a new digital differential analyser', *Simulation*, **24**, No. 4, pp. 97–102, April 1975.
4. Forbes, G. F., *The System Analyser*, **1**, No. 1, p. 11 (footnote), 1961.

7 Appendix 1: Serial arithmetic

Serial arithmetic is particularly simple and cheap to realize. The data are transmitted least significant bit first and normally the system operates in a strictly synchronous mode. As in any binary system, multiplication or division by 2 is achieved by shifting left or right once and in a serial system this can be done during the transmission of data between registers by simply inserting a D-flip-flop in the transmission path for multiplication (bringing about a shift in the arrival time of one clock period) or inserting a pulse at the end to cause an extra right shift.

Multiplication and division by 2, 4, 8, 16, etc., is therefore particularly attractive, although multiplication by any integer is not difficult, e.g.

$$y(t) \times 23 = y(t) \times (16+4+2+1)$$

producing the schematic diagram of Fig. 11, in which the symbol D stands for D-flip-flop.

Division, on the other hand, is only simple in general when the divider is a power of 2. Hence the preference for certain algorithms.

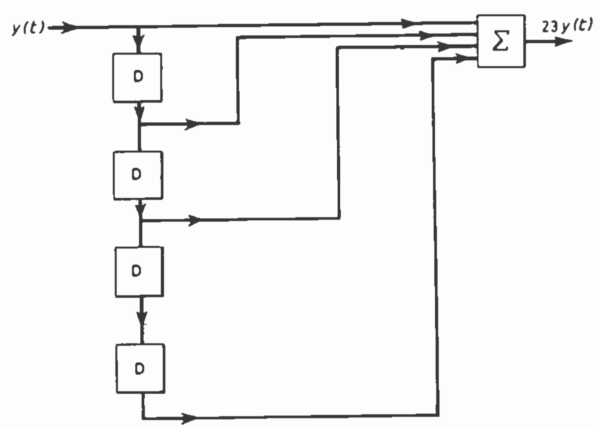


Fig. 11. A schematic diagram to illustrate the multiplication of $y(t)$ by an integer, 23, using D-flip-flops.

8 Appendix 2

8.1 Multiplier errors

Consider a multiplier with first-order prediction on both inputs x and y, the output being z.

The relationship between output and inputs is that

$$z(t+h)_p = x(t+h)_p * y(t+h)_p$$
$$= (2x(t)-x(t-h)) * (2y(t)-y(t-h))$$

Writing x, y for $x(t)$, $y(t)$ and using the Taylor expansion:

$$z(t+h)_p = (2x-(x-hx'+\tfrac{1}{2}h^2x''-\ldots)) * (2y-(y-hy'+\tfrac{1}{2}h^2y''-\ldots))$$
$$= xy+h(x'y+y'x)+ +h^2(x'y'-\tfrac{1}{2}x''y-\tfrac{1}{2}y''x)+\ldots \quad (30)$$

The true value of $z(t+h)$ is $x(t+h) * y(t+h)$ giving

$$z(t+h) = xy+h(x'y+y'x)+ +h^2(x'y'+\tfrac{1}{2}x''y+\tfrac{1}{2}y''x)+\ldots \quad (31)$$

Subtracting equation (30) from (31) gives the error:

$$E = h^2(x''y+y''x) \quad (32)$$

One could predict on the output of the multiplier instead of on the inputs, so that

$$z(t+h)_p = 2z(t)-z(t-h)$$
$$= 2z-(z-hz'+\tfrac{1}{2}h^2z''-\ldots)$$
$$= xy+h(x'y+y'x)+ +h^2(-x'y'-\tfrac{1}{2}xy''-\tfrac{1}{2}yx'')+\ldots \quad (33)$$

Subtracting equation (31) from (33) gives

$$E = h^2(x''y+y''x+2x'y') \quad (34)$$

Comparison of equations (32) and (34) indicates that input prediction is superior.

In a similar way one can show that the error with second-order prediction on the inputs is

$$E = h^3(x'''y+y'''x)$$

8.2 Divider errors

The inputs to the divider are x and y, the output is $z = x/y$.

The true value of $z(t+h)$ is given by

$$z(t+h) = \frac{x(t+h)}{y(t+h)} = \frac{x+hx'+\frac{h^2}{2}x''+\ldots}{y+hy'+\frac{h^2}{2}y''+\ldots} = \frac{x+a_x}{y+a_y}$$

where $x = x(t)$

$$a_x = hx'+\tfrac{1}{2}h^2x''+\ldots$$
$$a_y = hy'+\tfrac{1}{2}h^2y''+\ldots$$

Therefore

$$z(t+h) = (x+a_x)(y+a_y)^{-1}$$

But, using the binomial expansion

$$(y+a_y)^{-1} = y^{-1}-a_yy^{-2}+a_y^2y^{-3}-\ldots$$

Hence

$$z(t+h) = (x+a_x)(y^{-1}-a_yy^{-2}+a_y^2y^{-3}-\ldots)$$
$$= xy^{-1}-xa_yy^{-2}+xa_y^2y^{-3}+ +a_xy^{-1}-a_xa_yy^{-2}+\ldots \quad (35)$$

Neglecting terms in h^3 and above

$$-xa_yy^{-2} = -hy'xy^{-2}-\tfrac{1}{2}h^2y''xy^{-2}$$
$$xa_y^2y^{-3} = h^2xy^{-3}(y')^2$$
$$a_xy^{-1} = hx'y^{-1}+\tfrac{1}{2}h^2x''y^{-1}$$
$$-a_xa_yy^{-2} = -h^2x'y'y^{-2}$$

and equation (35) may be written

$$z(t+h) = xy^{-1}+h(x'y^{-1}-y'xy^{-2})+ +h^2(xy^{-3}(y')^2+x'y'y^{-2}+ +\tfrac{1}{2}x''y^{-1}-\tfrac{1}{2}y''xy^{-2}) \quad (36)$$

Using prediction (of order one) on the inputs:

$$z(t+h)_p = \frac{x(t+h)_p}{y(t+h)_p} \quad \frac{2x(t)-x(t-h)}{2y(t)-y(t-h)}$$
$$= \frac{2x-(x-hx'+\frac{1}{2}h^2x''-\ldots)}{2y-(y-hy'+\frac{1}{2}h^2y''-\ldots)}$$
$$= \frac{x+hx'-\frac{1}{2}h^2x''+\ldots}{y+hy'-\frac{1}{2}h^2y''+\ldots}$$
$$= \frac{x+b_x}{y+b_y}$$

where

$$b_x = hx'-\tfrac{1}{2}h^2x''+\ldots$$
$$b_y = hy'-\tfrac{1}{2}h^2y''+\ldots$$

Therefore

$$z(t+h)_p = (x+b_x)(y+b_y)^{-1}$$
$$= (x+b_x)(y^{-1}-b_yy^{-2}+b_y^2y^{-3}-\ldots)$$
$$= xy^{-1}-xb_yy^{-2}+xb_y^2y^{-3}+ +b_xy^{-1}-b_xb_yy^{-2}+\ldots \quad (37)$$

Neglecting terms in h^3 and above

$$-xb_yy^{-2} = -hy'xy^{-2}+\tfrac{1}{2}h^2y''xy^{-2}$$
$$xb_y^2y^{-3} = h^2(y')^2xy^{-3}$$
$$b_xy^{-1} = hx'y^{-1}-\tfrac{1}{2}h^2x''y^{-1}$$
$$-b_xb_yy^{-2} = -h^2x'y'y^{-2}$$

and equation (37) may be written

$$z(t+h)_p = xy^{-1}+h(x'y^{-1}-y'xy^{-2})+ +h^2((y')^2xy^{-3}-x'y'y^{-2}+ +\tfrac{1}{2}y''xy^{-2}-\tfrac{1}{2}x''y^{-1}) \quad (38)$$

The error using input prediction is given by subtracting equation (36) from (38):

$$E = h^2(x''y^{-1}-y''xy^{-2})$$
$$= \frac{h^2}{y^2}(x''y-y''x) \quad (39)$$

Using prediction on the output instead of the inputs:

$$z(t+h)_p = 2z(t)-z(t-h)$$
$$= 2z-\frac{x(t-h)}{y(t-h)}$$
$$= 2z-(x-b_x)(y-b_y)^{-1}$$

where b_x and b_y are defined as above

Therefore

$$z(t+h)_p = 2z-(x-b_x)(y^{-1}+b_y y^{-2}+b_y^2 y^{-3}+\ldots)$$
$$= 2z-xy^{-1}-xb_y y^{-2}-xb_y^2 y^{-3}+$$
$$+b_x y^{-1}+b_x b_y y^{-2}+\ldots$$
$$-xb_y y^{-2} = -hy'xy^{-2}+\tfrac{1}{2}h^2 y''xy^{-2}$$
$$-xb_y^2 y^{-3} = -h^2(y')^2 xy^{-3}$$
$$b_x y^{-1} = hx'y^{-1}-\tfrac{1}{2}h^2 x''y^{-1}$$
$$b_x b_y y^{-2} = h^2 x'y'y^{-2}$$

so that

$$z(t+h)_p = xy^{-1}+h(x'y^{-1}-y'xy^{-2})+$$
$$+h^2(x'y'y^{-2}-(y')^2 xy^{-3}+$$
$$+\tfrac{1}{2}y''xy^{-2}-\tfrac{1}{2}x''y^{-1}) \quad (40)$$

The error using output prediction is therefore, subtracting equation (36) from (40):

$$E = \frac{h^2}{y^2}\left((x''y-y''x)+\frac{2y'}{y}(y'x-x'y)\right) \quad (41)$$

The existence of the extra term in equation (41), compared to equation (39) suggests that in general prediction on the inputs is again superior.

9 Appendix 3

9.1 Prediction routines

$$y(t+h) = y+hy'+\tfrac{1}{2}h^2 y''+\frac{1}{3!}h^2 y'''+\ldots \quad (42)$$
$$= a_0 y_0+a_1 y_1-a_2 y_2+a_3 y_3+\ldots \quad (43)$$

where

$$y_0 = y(t) = y$$
$$y_1 = y(t-h)$$
$$y_2 = y(t-2h), \quad \text{etc.}$$

and $a_0, a_1, a_2, \ldots$ are the coefficients of the prediction routine. So equation (43) becomes

$$y(t+h) = a_0 y+$$
$$+a_1\left(y-hy'+\tfrac{1}{2}h^2 y''-\frac{1}{3!}h^3 y'''+\ldots\right)+$$
$$+a_2\left(y-2hy'+\tfrac{1}{2}4h^2 y''-\frac{1}{3!}8h^3 y'''+\ldots\right)+$$
$$+a_3\left(y-3hy'+\tfrac{1}{2}9h^2 y''-\frac{1}{3!}27h^3 y'''+\ldots\right)+$$
$$+\ldots \quad (44)$$

To construct a predictor of any chosen order we simply equate terms in equations (42) and (44). For instance, a first-order predictor gives

$$y+hy'+\tfrac{1}{2}h^2 y''+\ldots \simeq a_0 y+a_1(y-hy'+\ldots)$$

and equating terms in y and y'

$$a_0+a_1 = 1$$

and

$$a_1 = -1$$

Therefore

$$a_0 = 2$$

Thus

$$y(t+h) = 2y(t)-y(t-h)$$

The second-order predictor:

$$y+hy'+\tfrac{1}{2}h^2 y'' \simeq a_0 y_0+a_1 y_1+a_2 y_2$$
$$= y(a_0+a_1+a_2)+hy'(-a_1-2a_2)+$$
$$+\tfrac{1}{2}h^2 y''(a_1+4a_2)$$

Hence

$$a_0+a_1+a_2 = 1$$
$$-a_1-2a_2 = 1$$
$$a_1+4a_2 = 1$$

Therefore

$$a_2 = 1, \qquad a_1 = - \;, \qquad a_0 = +3$$

i.e.

$$y(t+h) = 3y(t)-3y(t-h)+y(t-2h)$$

9.2 Integration routines

$$z(t) = \int_{t-h}^{t} y(t)\,dt = Y(t)-Y(t-h)$$

where

$$Y(t) = \int y(t)\,dt$$
$$= Y-\left(Y-hY'+\tfrac{1}{2}h^2 Y''-\frac{1}{3!}h^3 Y'''+\ldots\right)$$
$$= hy-\tfrac{1}{2}h^2 y'+\frac{1}{3!}h^3 y''-\frac{1}{4!}h^4 y'''+\ldots$$

since $Y' = y$, $Y'' = y'$, etc.

$$= h\left(y-\tfrac{1}{2}.hy'+\tfrac{1}{3}.\tfrac{1}{2}h^2 y''-\tfrac{1}{4}.\frac{1}{3!}h^3 y'''+\ldots\right) \quad (45)$$
$$= h(b_0 y+b_1 y_1+b_2 y_2+\ldots) \quad (46)$$

where b_0, b_1, etc., are the coefficients of the integration routine.

For a first-order integrator

$$h(y-\tfrac{1}{2}hy'+\tfrac{1}{3}.\tfrac{1}{2}h^2 y''-\ldots)$$
$$= h(b_0 y+b_1(y-hy'+\tfrac{1}{2}h^2 y''-\ldots))$$

Equating terms

$$b_0+b_1 = 1$$

and

$$b_1 = \tfrac{1}{2}$$

Therefore

$$b_0 = \tfrac{1}{2}$$

So

$$z(t) = h(\tfrac{1}{2}y(t)+\tfrac{1}{2}y(t-h))$$

Similarly, a second-order integrator requires the algorithm

$$z(t) = \frac{h}{12}(5y(t)+8y(t-h)-y(t-2h))$$

9.3 Combined prediction and integration routines

These may be derived in a similar fashion, from the equation:

$$z(t+h) = \int_t^{t+h} y(t)\,\mathrm{d}t \qquad (47)$$

resulting in

$$z(t+h) = h\left(y+\tfrac{1}{2}.hy'+\tfrac{1}{3}.\tfrac{1}{2}h^2y''+\tfrac{1}{4}.\frac{h^3}{3!}y'''+\ldots\right) \qquad (48)$$

$$= h(c_0y_0+c_1y_1+c_2y_2+\ldots) \qquad (49)$$

Carrying on as before we obtain the first-order routine

$$z(t+h) = \frac{h}{2}(3y(t)-y(t-h))$$

and the second-order routine

$$z(t+h) = \frac{h}{12}(23y(t)-16y(t-h)+5y(t-2h))$$

etc.

9.4 It may strike the reader that the procedures explained above lend themselves to matrix formulation.

Let T be the $n\times n$ matrix, defined for any n:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \ldots \\ 1 & -1 & 1 & -1 & 1 & \ldots \\ 1 & -2 & 4 & -8 & 16 & \ldots \\ 1 & -3 & 9 & -27 & 81 & \ldots \\ 1 & -4 & 16 & -64 & 256 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

Let C be the $(1\times n)$ row matrix

$$C = [a_0 \quad a_1 \quad a_2 \quad a_3 \quad \ldots \quad a_{n-1}]$$

where $a_0, a_1, \ldots, a_n$ are the coefficients in the predictor algorithm as defined above.

Let D be the $(1\times n)$ row matrix

$$D = [1 \quad 1 \quad 1 \quad \ldots \quad 1]$$

To determine the elements of the C matrix, the unknowns, we must solve the equation

$$CT = D$$

i.e.

$$C = DT^{-1}$$

in which $n = k+1$ for a kth-order predictor.

To find the coefficients of a kth-order integration algorithm then:

$$C = [b_0 \quad b_1 \quad b_2 \quad \ldots \quad b_k]$$

and

$$D = [1 \quad -\tfrac{1}{2} \quad \tfrac{1}{3} \quad -\tfrac{1}{4} \quad \tfrac{1}{5} \quad \ldots] \text{ to } k+1 \text{ terms}$$

while for a kth-order predictor–integrator (the $k.k$ integrator)

$$C = [c_0 \quad c_1 \quad c_2 \quad \ldots \quad c_k]$$

and

$$D = [1 \quad \tfrac{1}{2} \quad \tfrac{1}{3} \quad \tfrac{1}{4} \quad \tfrac{1}{5} \quad \ldots] \text{ to } k+1 \text{ terms}$$

For example, the second-order predictor gives

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 1 \\ 1 & -2 & 4 \end{bmatrix} \quad \text{and} \quad T^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \tfrac{3}{2} & -2 & \tfrac{1}{2} \\ \tfrac{1}{2} & -1 & \tfrac{1}{2} \end{bmatrix}$$

Therefore

$$C = [1 \quad 1 \quad 1][T^{-1} =][3 \quad -3 \quad 1]$$

The second-order integrator gives

$$C = [1 \quad -\tfrac{1}{2} \quad +\tfrac{1}{3}][T^{-1}] = [\tfrac{5}{12} \quad \tfrac{8}{12} \quad -\tfrac{1}{12}]$$

and the 2.2 integrator gives

$$C = [1 \quad \tfrac{1}{2} \quad \tfrac{1}{3}][T^{-1}] = [\tfrac{23}{12} \quad -\tfrac{16}{12} \quad \tfrac{5}{12}]$$

Manuscript first received by the Institution on 14th July 1975 and in final form on 9th January 1976. (Paper No. 1745/Comp. 177).

The Authors

William Forsythe graduated from The Queen's University of Belfast in 1960 and worked for four years for Hawker Siddeley Dynamics (Guided Weapons Division) before taking up his present appointment as a Lecturer in the Department of Electronic and Electrical Engineering at Loughborough University. His main research interests are in control engineering and associated areas.

Stephen Houseman was an apprentice technician with English Electric at Preston until 1968 when he went to Loughborough University to read electrical engineering, graduating in 1971. For the next three years he continued as a research student investigating parallel processing problems in the Department of Electronic and Electrical Engineering, and since 1974 he has been on the staff of IBM, Hursley Park.

UDC 621.385.3.032.262.3: 621.373.42: 621.365.55

Controlled r.f. power generation using magnetically-beamed triodes

R. W. H. WHITTLE*
and
J. J. BEHENNA, D.F.H., C.Eng., M.I.E.E.†

Based on a paper presented at the Components and Circuits Group Colloquium on H.F. Heating Circuits and Techniques held in London on 9th October 1974

SUMMARY

The magnetically-beamed triode has a structure radically different from a conventional industrial triode. The relatively small proportion of the current which is intercepted by the control electrode (gate), even when it is driven to a high positive potential, leads to useful design properties and novel applications circuits.

The cool gate needs no special coating to inhibit electron emission and oxide-coated cathodes are used to complete a robust structure and reduce ancillary heater power.

Low gate current results in low bias loss and low oscillator drive power. The latter permits the introduction of a new form of oscillator output power control by limiting the gate voltage swing by a catching-diode in conjunction with a solid-state regulator. Design parameters are considered for this unit which is less costly than alternative systems. Oscillator power may be controlled from maximum down to 10% by the swing of only a few volts at the regulator input. Provision may be made for external feedback loops in addition to a manual presetting potentiometer. Examples are quoted for the stabilization of input current or r.f. output voltage.

*ITT Components Group, Electron Device Division, Standard Telephones and Cables Ltd., Brixham Road, Paignton, Devon TQ4 7BE.
† Formerly with ITT Components Group Europe.

1 Introduction

In the infancy of r.f. industrial heating, valves were used which had been specifically designed for use in telecommunication transmitters. Although some of these valves were, with care, used successfully in early equipments, it soon became apparent that their characteristics and mechanical construction resulted in quite severe limitations in their application in this new field. The primary reason for this is that transmitters are in general subjected to a constant electrical load which allows optimum matching to be obtained between the valve and the circuit to which the power is delivered, whereas for industrial oscillators the situation is quite different. Here there can be large differences in load impedance, for example, when a ferric load goes through the Curie point, and in consequence both the power delivered by the valve and the power dissipated on its electrodes are subject to considerable variation. This is particularly true for the control grid where the dissipation under low load conditions can be several times the dissipation under full load. This often resulted in the maximum allowable grid dissipation being exceeded when transmitting valves were used, leading to high grid emission, 'run-away' conditions, and consequent valve failure. In the mid-1950s, valves were specifically designed to operate with adequate safety factors under the onerous conditions met in industrial heating applications. A number of different approaches were made to solve this problem, such as optimizing the amplification factor[1] to minimize the variation in output power and grid dissipation with changing load impedance; alternatively lowering the amplification factor[2,3] to design in a large grid safety factor.

Though these improvements were significant, the construction of tubes remained conventional with the fundamental fragility of grids and thoriated tungsten filaments. The introduction of a magnetic beaming,[5] which dramatically improves the division of current between anode and grid (now called gate) represents the most significant change in the last quarter-century. Figure 1 compares the grid/cathode structures of a mesh industrial triode with that of a magnetically-beamed triode.

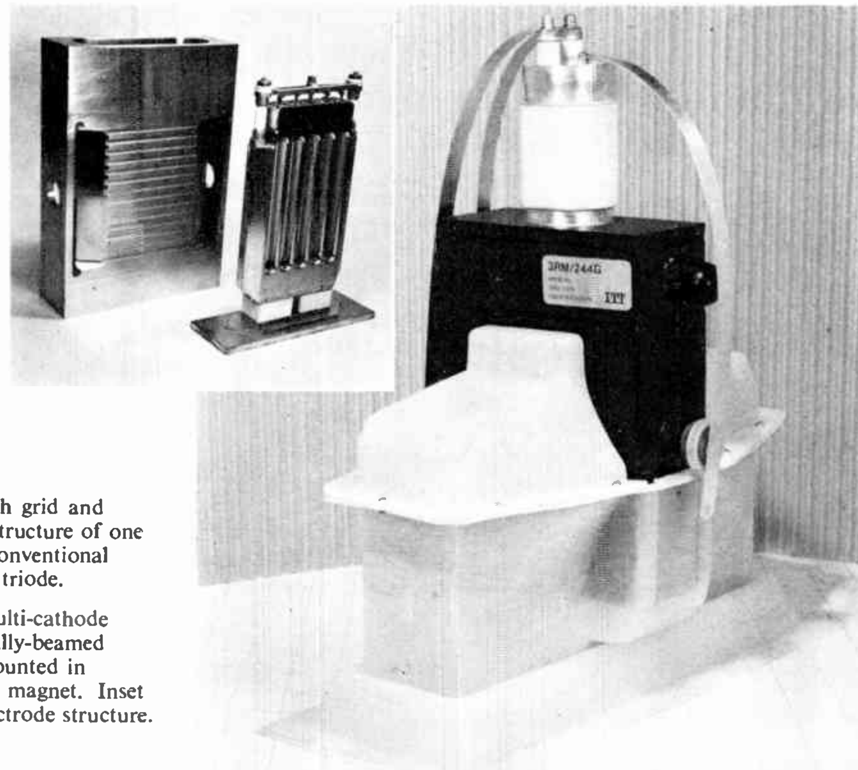
2 Principle of Operation of Magnetically-Beamed Valves

In the magnetically-beamed triode, the problems referred to earlier have been overcome by replacing the normal wire grid by a gate electrode. Electronic emission from the cathode is focused by a magnetic field into a rectangular beam of current, which is collected by the anode. Modulation of the beam is performed by the gate, which is equivalent to the grid of a conventional triode. The combined effect of the magnetic field and the perpendicular components of electrostatic field resulting from the anode and gate potentials causes emitted electrons to spiral towards the anode.

The radius of curvature R of this helix is dependent on the magnitudes of the electrostatic and magnetic fields, in accordance with the relation

Fig. 1.
(left) Mesh grid and filament structure of one form of conventional industrial triode.

(right) Multi-cathode magnetically-beamed triode mounted in shrouded magnet. Inset shows electrode structure.



$$R \propto \sqrt{V_r}/B$$

where V_r = voltage component perpendicular to the magnetic field,

B = flux density.

The pitch P is given by

$$P \propto \sqrt{V_p}/B$$

where V_p = voltage component parallel to the magnetic field.

The intercepted gate current is an inverse function of magnetic field but is relatively independent for field strengths well above the 'knee' region (Fig. 2).

As with conventional triodes, tube characteristics may be described in terms of mutual conductance and amplification factor, and are determined by geometric relationships, particularly gate dimensions and gate-cathode spacing: dimensions m, n, d in Fig. 3. The present range of m.b. triodes has an amplification factor of about 25 which is near optimum[1] for industrial oscillators.

3 Construction

The very low gate dissipation of a magnetically-beamed triode when operated as an induction heating oscillator, typically one-sixth of that on the grid of an equivalent conventional triode, allows a wide choice of materials for the gate construction. In the range of valves described here, plain copper replaces the molybdenum and multi-layer anti-emissive coatings used for conventional grids. The gates are conduction cooled, through beryllia† blocks, to the water-cooled anodes. In conjunction with the low dissipation, this allows gate operating temperatures as low as 250°C to be attained,

† The latest design has eliminated the need for this material.

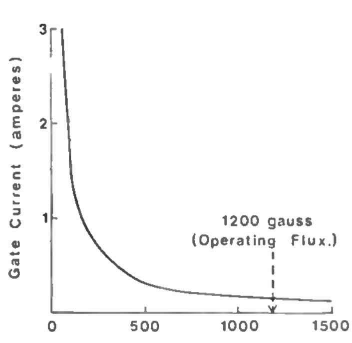


Fig. 2. Intercepted gate current vs. magnetic field (3RM/245G).

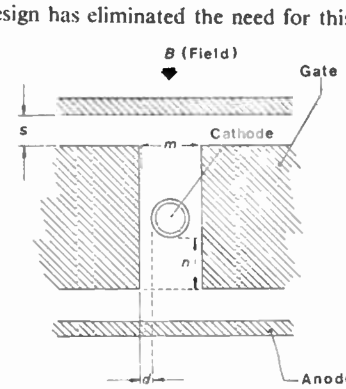


Fig. 3. Electrode and magnetic field configuration.

Table 1

		M	C	M	C	M	C
Valve Type		3RM/189G	3R/199E	3RM/244G	3R/252E	3RM/245G	3R/262E
Anode voltage	kV	6·5	6·5	9·0	9·0	7·0	7·0
Anode current	A	1·8	1·8	5·7	5·7	6·8	6·8
Gate/grid current	mA	25	500	70	475	70	500
Gate/grid resistor	Ω	30 000	900	13 500	1650	10 000	1700
Gate/grid dissipation	W	20	150	45	130	40	130
Gate/grid resistor dissipation	W	20	200	65	220	49	430
Cathode heating power	W	79	403	176	1000	209	2400
Output power (anode)	kW	8·1	8·6	35	35	35	33
Drive power (feedback)	W	37	350	110	350	90	580
Net output	kW	8	8·2	34·9	34·6	34·9	32·4
Efficiency (anode)	%	68·0	70·0	68·0	68·0	73·0	68·0
(overall		67·9	67·7	67·8	66·0	72·9	64·8

M—magnetically beamed triode C—conventional triode

typically about 800°C below the temperature of grids in conventional valves. This means that anti-emissive coatings are no longer required.

A further important advantage results from the low operating temperature of the gate in that it allows the use of oxide-coated cathodes in place of the thoriated tungsten filaments used in conventional triodes, with a consequent saving in heater power of about 80%. Oxide cathodes cannot be used with grids operating at high temperature because any barium reaching the grid by evaporation from the cathode would cause excessive electron emission from the grid. The direction of this current would be in opposition to the normal grid current flow and would effectively reduce the negative bias. This is because the bias is derived entirely by the use of a grid bias resistor which, apart from being relatively cheap, effects some compensation by adjusting the bias against varying grid drive voltage. The loss of bias would lead, in turn, to increased anode and grid currents, higher grid temperature and electron emission and hence run-away conditions.

The drive power requirement for m.b. triodes is far below that required for a conventional valve. In the same proportion so are the gate dissipation and gate leak resistor dissipation. To achieve the required bias voltages the ohmic value of the leak resistor must be proportionally higher, but its physical size and cost are reduced roughly in proportion to dissipation.

Table 1 makes quantitative comparison between three m.b. triodes and their nearest equivalents (in terms of supply voltage and output power) among conventional triodes.

4 Oscillator Circuits

Magnetically beamed triodes are used in industrial oscillator circuits of conventional design such as Colpitts, or tuned anode/coupled grid (which is essentially a split-coil Hartley) and others. Since exactly the same voltage and power handling conditions prevail in the anode circuit, component values here are traditional. The differences occur in the control electrode circuit; as already mentioned, the leak resistor has a substantially higher value (typically 10 times greater). To avoid 'squegging' the gate coupling capacitor must be reduced by a corresponding factor. This presents no problem in increased reactive impedance since the drive current which it must pass is commensurately lower.

Figure 4 illustrates a typical tuned anode/coupled gate oscillator together with output power control by gate limiting circuit described in Section 3.3.

5 Output Power Control

The control in output power from an oscillator may be effected in several ways which divide into three main groups:

1. Control of the main supply voltage.
2. Change of load coupling away from that for optimum efficiency.
3. Control of the oscillator behaviour in the grid or gate circuit.

The first group sub-divides again into three classes:

1.1 The use of a variable transformer regulator. This is an expensive and bulky item limited to mechanical adjustment only.

1.2 Control of the e.h.t. power supply by delayed firing of thyratrons or, in modern technology, thyristors. These are high power (and relatively high cost) elements and are prone to the generation of spiky ripple on the e.h.t. voltage, and resultant troublesome feed-back into the electricity supply authority's distribution network.

1.3 Control of a series element between the e.h.t. supply and the oscillator circuit. This system is satisfactory at the expense of an extra power tube. Normally this is a saturable, tungsten filament diode but the m.b. triode itself offers easier electronic control, fail-safe features and economy. The m.b. tube has the advantages that the relatively high amplification factor at low anode currents makes the anode current at given bias level relatively independent of voltage drop: furthermore, the low gate currents permit the control voltage to remain steady in the positive region and within dissipation limits even at low anode voltages. As a 'fail-safe' feature the bias may be tripped out in the event of a detectable malfunction of the equipment to reduce the oscillator output power to a low level. By contrast, a low impedance conventional triode as a series regulator would normally operate in the negative bias region and loss of such bias would result in a dangerous, higher than normal current flow.

Considering the second group:

2.1 Power control by effectively mismatching the load to the oscillator is usually only achievable by mechanical means and leads to lower efficiency. Without reducing the e.h.t. voltage the only direction in which the load matching can be safely offset is by uncoupling, i.e. decreased effective load admittance. If this is done to limit the heating rate of a ferrous load below Curie point, the mismatch becomes even worse above that temperature just when tighter coupling is desirable. Overcoupling to a load results in higher anode current, lower efficiency and the risk of excessive dissipation in the valve anode.

The third grouping also sub-divides:

3.1 Interrupted cut-off technique

This was originally conceived for use with conventional tubes and is even easier to achieve with m.b. triodes because it is so much easier to develop a cut-off bias across their relatively high resistance grid (gate) leak component. The power control is effected by varying the mark space ratio in which oscillations are interrupted and the method does have the merit of running the oscillator at full efficiency during the ON periods. However, there is a serious disadvantage in that convenient interruption frequencies are in the audio frequency range, e.g. 50 (or 60) Hz and a corresponding low-frequency component becomes superimposed upon the radio frequency output of the oscillator. This can present an operator with a severe shock hazard if the work coil is inadvertently touched. Hence, the use of this system is inadvisable except in remote process applications.

3.2 Reduced drive coupling

This is limited in range because the oscillator efficiency is reduced and there is risk that oscillation will cease. Furthermore, such an arrangement would be mechanical and not readily adaptable to an electronic feedback system.

3.3 Variable gate loading techniques[6]

In early experiments with m.b. triodes, it was discovered that variation of the magnetic field (towards the knee region) produced a significant change in output power and that power control by a mechanical system of adjustable magnetic shunts was feasible. It was deduced that the fundamental effect of this alteration of magnetic field was change in gate current and that this could be effected electronically by a shunt element, either thermionic valve or semiconductor.

Initially shunt triodes and tetrodes were tried, but difficulty was experienced with screen grid dissipation or finding a triode of sufficiently low impedance for wide range power control. For low impedance a power transistor was more attractive as well as more fashionable. The use of a series connected (thermionic) diode to block the destructive reverse voltage led directly to the development of the gate-limiter system of power control which is the principal topic of the final Section of this paper.

6 Gate Limiting Power Control

The basic circuit is shown in the lower part of Fig. 4. A low-power, high-voltage diode D1 connects the gate of the m.b. power triode to a potential divider formed between resistor chain R1 (comprising a chain of about 100 kΩ per kilovolt e.h.t. voltage) and the power transistor shunt regulator. If the regulator is set in a turned-off condition, such that the potential at the diode cathode exceeds the peak of the m.b. triode gate drive signal, the diode is backed-biased throughout the whole r.f. cycle and the oscillator yields full power—the same as if no power controller were attached.

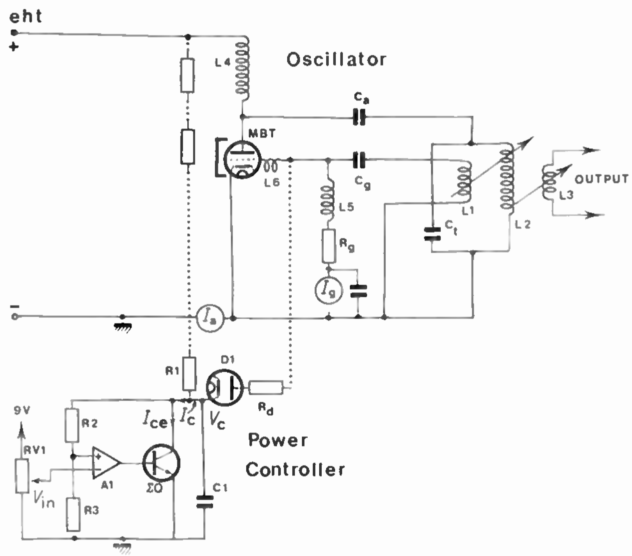


Fig. 4. Tuned anode/coupled gate oscillator with gate limiting circuit for output power control.

If the shunt regulator voltage is reduced until the diode begins to conduct at the crest of the drive waveform, the drive is reduced, the m.b. triode anode voltage swing is reduced, the oscillator output power is reduced but so, also, is the gate drive signal. Hence the oscillator settles to a new lower output power level. Thus, whilst the diode and regulator must have a dynamic impedance low compared with the drive source impedance of the oscillator (substantially the reactive impedance of capacitor C_g), they need clip only the crest of the new drive waveform rather than to load it at its original amplitude and this requires the controller to dissipate only a small power relative to that which it controls.

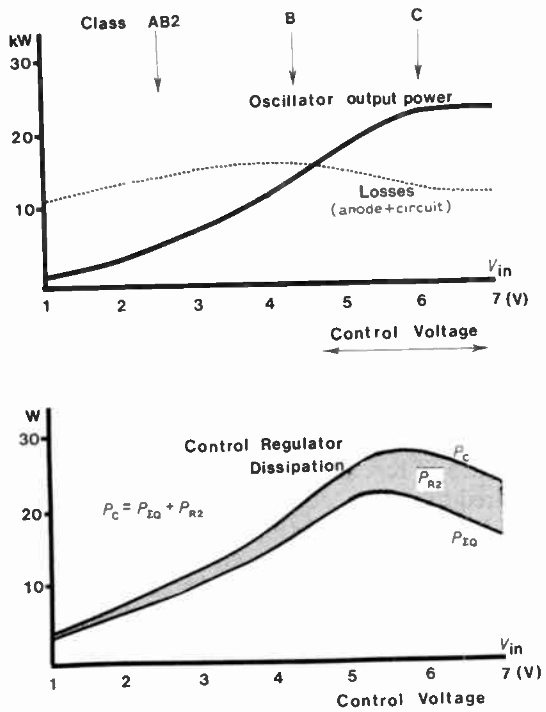


Fig. 5. Oscillator power and dissipation related to control setting.

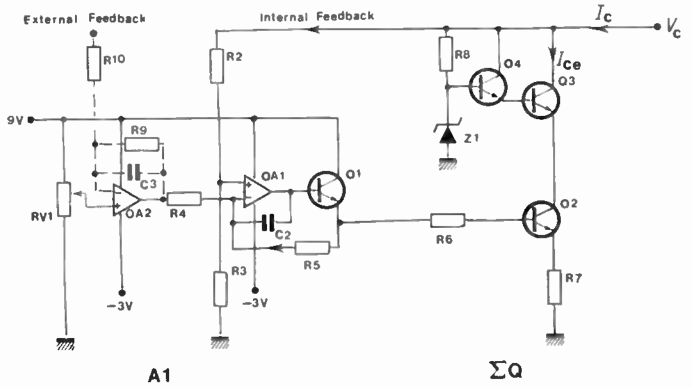


Fig. 6. Power control regulator.

The smoothness and range of such a power-controlled oscillator is typified by the output power curve in Fig. 5 which is plotted against control potentiometer voltage. It will be seen that the control regulator dissipation (P_c) is 3 orders smaller than the power controlled and is shared between the regulator transistor(s) $P_{\Sigma Q}$ (lower curve) and the feedback resistor P_{R2} (shaded area).†

In its simplest form, the shunt regulators may comprise a single power transistor with its base connected to the level-set potentiometer. However, in most applications it is desirable to insert an amplifier, with differential input to enable a negative feedback signal to be applied via divider R2/R3 (Fig. 4).

R2 is typically 100 kΩ and R3 pre-set to about 750 Ω; thus at a maximum regulator voltage of 900 V the feedback is about 7 V, which is within the range that can be balanced by the input from RV1 (either direct or via OA2 as in Fig. 6). Lower inputs from RV1 cause the feedback via R2/R3 to stabilize the regulator voltage at the lower levels of V_c for which RV1 has been set.

Capacitor C1 has a low reactive impedance at the oscillator frequency, removes ripple and thus makes the control regulator essentially a steady direct-current device.

The differential drive amplifier may comprise an operational amplifier OA1 (Fig. 6) plus a low-power transistor Q1 in emitter-followed connection to provide adequate current to the base of power transistor Q2 to turn it hard on for minimum oscillator power. Feedback loops (C2, R5) within this drive amplifier circuit limit overall gain to a suitable level (typically 10 dB set by ratio R5/R4) and provide stability. The reactance of C2 effectively shunts R5 and its low value (relative to R4 = 15 kΩ), at high frequencies restricts the frequency range to avoid response to any spurious signals.

The broken lines in Fig. 6 (A1) show how a second differential-input operational amplifier OA2 further splits the input options to feedback from an external sensor in addition to the manual level-set (RV1). The gain of OA2 is likewise set by the resistor ratio R9/R10 and the frequency response limited by the capacitor C3. The gain of OA2 coarsens the response to RV1 and its adjustment relative to the feedback signal is mentioned later.

7 Power Dissipation Ratings

Reference to Fig. 5 shows that oscillator power loss (mostly anode dissipation) and controller power loss both peak at intermediate power levels. The anode dissipation increases because the anode voltage swing is reduced and the anode current peak, though lower, occurs at a higher anode voltage. The relationships of voltages and currents at anode and grid (gate) of a power oscillator tube are usually analysed by reference to the tube's constant-current characteristics. Both anode and grid voltages are sine-wave and, for a resistively-loaded oscillator, their relationship is a straight line with its centre on the ordinate through $V_a = V_s$, the anode supply voltage. In Fig. 7, only one half of each operating line is shown. (In class B or C no current flows during the other half-cycle and it does not enter into the calculation of mean or peak-fundamental currents. Such calculations are discussed more fully in Appendix 1.)

At full output it is conventional and sensible to set the conditions to achieve the best efficiency in class C. As the gate voltage swing is limited, the gate current and, for fixed leak resistor, the gate bias $V_g = I_g R_g$ is reduced; hence the (half) operating line moves as shown through class B to AB2 with reduced input current but also reduced efficiency.

At low power levels the loss in efficiency is not serious because the input power is low. Many applications will use the oscillator near full output and considerations for high reliability in an industrial environment demands that adequate anode cooling be provided. Excellent margin is available in actual tube ratings.

† $P_{\Sigma Q}$ is the total power dissipated in transistors ΣQ in Fig. 4 and P_{R2} is dissipated in R2.

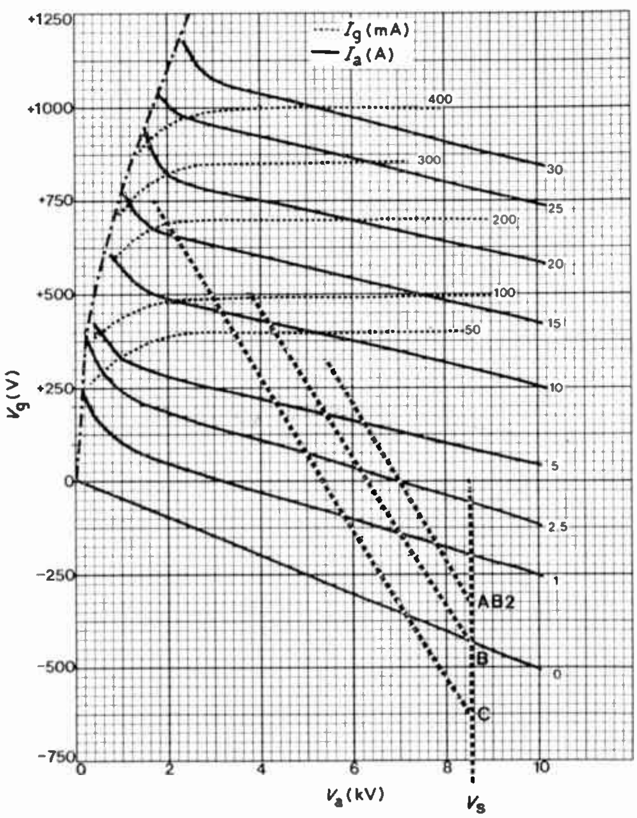


Fig. 7. Typical constant current characteristics (3RM/244G).

By the same token, the control regulator must also be reliable. Care must be taken with both dissipation and voltage ratings of the power transistor(s) and voltage/current plots of this part of the circuit need to be made over the full range of control of the oscillator with the various degrees of drive coupling and work load impedance likely to be experienced. If this V_c/I_c plot is made on log-log graph paper the constant power curves are straight lines and may be readily added as an additional graticule; likewise the effective total impedance, though this is mostly of academic interest.

Figure 8 shows such a set of curves (thick lines) for a range of oscillator loading. (Tube type 3RM/244G operated at 8·5 kV anode voltage.) These dynamic curves intersect the (broken) static curves at the corresponding input control voltage, V_{in}, applied to resistor R4 (Fig. 6). Each dynamic curve corresponds to a particular load coupling. The left-hand curve corresponds to optimum coupling for maximum performance.

The static curves may be plotted by connecting the regulator (separated from a power oscillator) across a d.c. power supply plus a series load resistor. The independence of regulator voltage relative to the current it is required to pass demonstrates the effectiveness of the feedback loop via R2, R3 (Figs. 4 and 6). The locus of the 'knee' at the low current end of the static curves corresponds to an impedance marginally smaller than R2. It would be exactly equal to this value if the transistor chain ΣQ were to be cut off completely and R8 were infinite rather than the same order as R2.

Reduced gate drive impedance moves a given dynamic curve in the same direction as reduced load admittance and if the gate coupling capacitor (C_g in Fig. 4) is larger than necessary it requires the transistor regulator to have an unnecessarily high dissipation rating. In this matter special design care needs to be exercised with respect to feedback coupling in a Colpitts oscillator circuit.

If the dissipation margin is insufficient with a single power transistor, a second transistor must be provided. Shunt connection is readily achievable with low value emitter resistors† to balance the current. Alternatively, if the voltage rating is insufficient, series connection must

† Positioned as R7 in Fig. 6.

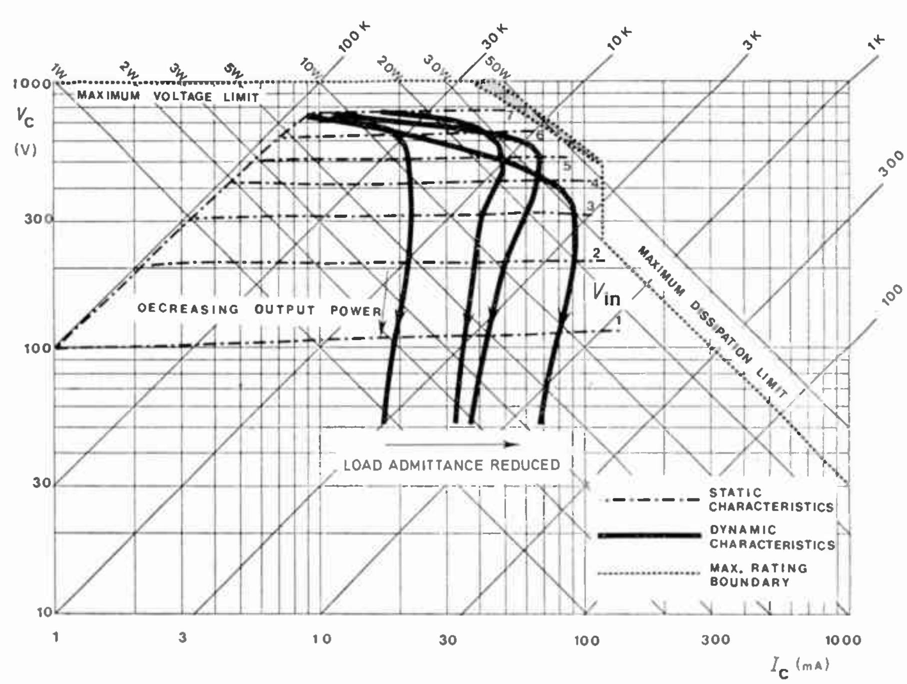


Fig. 8. V_c/I_c plot for range of oscillator loadings for 3RM/244G.

be used. Equal sharing of voltage by a ladder of resistors linking the transistor bases presents problems in achieving the extremes of turn-on and turn-off. A more effective arrangement is shown in Fig. 6. When Q2 is turned on hard, the upper transistor Q3 is also turned on hard by current through R8 amplified by Q4. (Q4 must also have a high voltage rating but its dissipation is considerably lower than Q3.) The voltage across Q2 is low and as Q3 and Q4 are in emitter-follower connection, the voltage at the base of Q4 is not much higher and well below the conduction level of Zener diode chain Z1.

As Q2 is turned off by the level set (or external feedback), it takes most of the voltage V_c until the Zener diode Z1 voltage is reached. Then the V_{CE} of Q2 is held at that level and further voltage rise is taken up by Q3 (and Q4): much of the current through R8 is now directed from the base of Q4 through the Zener chain Z1.

By careful choice of Z1 Zener chain voltage and series/parallel combinations of Q2 and Q3, the voltage/dissipation rating boundary curve of the total unit can be tailored to provide a good safety margin relative to the most onerous V/I curve plotted empirically. The rating boundary shown in Fig. 8 assumes a Zener chain (Z1) voltage of 250 V and that the V_{CE} rating of Q2 exceeds this value: also that the V_{CE} rating of Q3 and Q4 is 750 V and that the maximum dissipation of Q2 and Q3 is 30 W each at the maximum heat sink temperature to be encountered.

The shaded area of the boundary curve is the power dissipated by feedback resistor R2. Formulae defining the ratings boundary are given in Appendix 2.

8 Power Controller Applications

Some industrial heating applications merely need a manual control or presetting by the operator: other processes benefit, or indeed may only be successful if the change of power level is rapid, thus needing automatic feedback or programming. The low voltage and power level at the input end of the regulator driver circuits make them ideal for coupling to solid-state logic circuits for programming oscillator output power.

If it is desirable to maintain constant input power to the oscillator, the anode current (or, more readily, the cathode current which is nearly the same) may be monitored as a voltage by insertion of a low-ohmic resistor plus integrating circuit to obtain a mean level for feedback to the auto-control input (Fig. 9(a)).

Perhaps, more usefully, the r.f. output voltage can be kept constant by monitoring its level converted to low voltage d.c. via a capacitive potential-divider, detector diode and smoothing circuit (Fig. 9(b)).

If component values, including those in the regulator amplifier, are chosen for reasonable frequency response up to 300 Hz, then the ripple at that frequency from the usual 3-phase, full-wave rectified, e.h.t. power supply can be suppressed significantly.

Feedback may be from a point even nearer the load in the form of an electronic temperature sensor with an output voltage which may be suitably amplified for presentation to programmed logic control, or, for simple

(a) Automatic current regulation.

(b) Automatic r.f. voltage regulation.

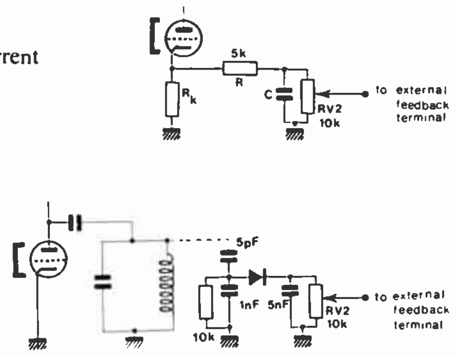


Fig. 9.

stabilization of load temperature, directly to the external feedback terminal of the regulator.

For practical realizations, it is convenient to develop the net external feedback voltage across a 10 kΩ potentiometer RV2, the sliding contact of which is connected to the regulator input. Since this input level acts in conjunction with, and in the reverse sense to, the manual preset (RV1 in Fig. 6), it is a useful setting-up technique to start with RV2 output at zero, set the oscillator output level with RV1 and finally adjust RV2 and RV1 alternately to keep the required output level whilst introducing sufficient feedback for automatic control of the process or stabilization of work-coil voltage or work-load temperature, as the case may be.

9 Conclusions

A new concept of industrial oscillator valve design and output power control has now been substantially proved in laboratory test and workshop floor experience in a number of industrial heating equipments of various designs and powers ranging from 6 to 30 kW.

These new methods of power control open the way for more refined industrial heat treatment techniques in addition to substantial cost saving.

10 Acknowledgments

The authors are indebted to a number of helpers, especially to Mr. G. H. Phipps who was involved in the detail of the valve engineering, to Mr. B. G. Kennett who pioneered the gate-limiting system of power-control, and to Mr. A. E. Miller who contributed significantly to the detail of the circuit design and carried through the extensive proving tests. Thanks are also due to the Management of ITT Components Group, Europe for permission to publish.

11 References

1. Dorgelo, E. G., 'Output and load resistance of oscillating triodes in r.f. heating generators', *Electronic Applications Bull.*, **18**, No. 1, pp. 19–26, 1957.
2. Pohl, W. J., 'The design and operation of high power triodes for radio frequency heating', *Proc. Instn Elect. Engrs*, **104**, part B No. 16, pp. 410–6, July 1957.
3. Behenna, J. J., 'Ceramic-insulated vacuum tubes for very high frequency industrial heating', *Electrical Communication*, **38**, No. 3, pp. 396–406, 1963.

4. Green, B., 'Quick calculations for r.f. amplifiers', *Electronic Design*, **11**, pp. 52–4, 21st June 1963.

5. Behenna, J. J., Kennett, B. G. and Phipps, G. H., '25 kW Magnetically-focussed Triode', Proc. Electronic Components Conference, RECMF, London, May 1971.

6. ITT Application Note: 6312/649E. 'Magnetically-focussed Triodes', 1973/74.

12 Appendix 1: Oscillator performance analysis[4]

The general-case locus of the V_a/V_g relationship is an ellipse which virtually collapses to a straight line AGA' for a resistively-loaded oscillator (Fig. 10). (Significantly reactive loads can cause serious tube dissipation problems.) In Fig. 10, G is at the centre of the line and has co-ordinates

x = anode supply voltage V_s

y = grid bias voltage $V_{gb} = -I_g R_g$

where I_g is the mean grid (gate) current and R_g is the leak resistor shown in Fig. 4.

As the grid voltage excursion is reduced by the gate limiting system of power control, the gate current falls and the reduction of $I_g R_g$ causes the movement of the point G to a less negative level. Point G_1 corresponds to the condition B in Fig. 7 and G_2 corresponds to condition AB2.

The locus of A, A_1, A_2 is not necessarily a straight line but it is approximately so in the present example. Line length AA' reduces and finally vanishes at $V_a = V_s$, $V_g = 0$. In practice oscillation would cease before this point was reached. This situation does not occur with tubes of present design if the gate catching potential V_c is not taken below zero.

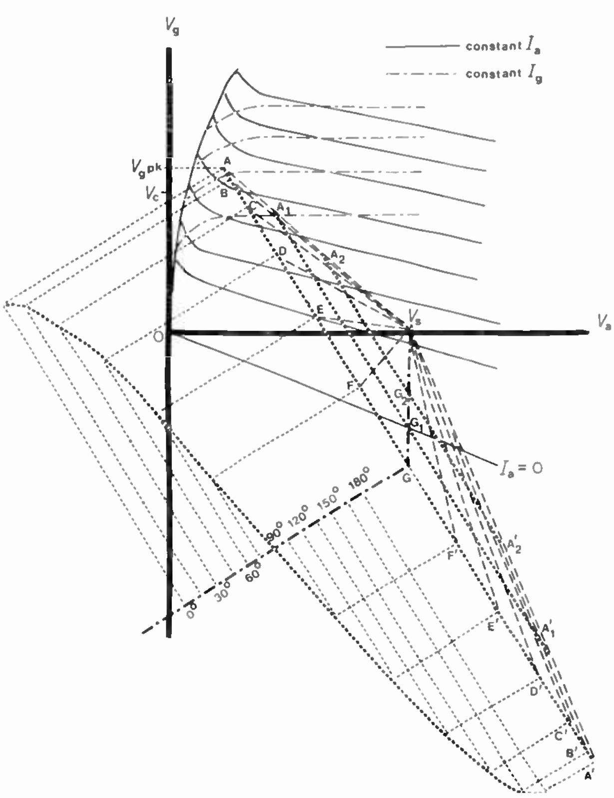


Fig. 10. Locus of V_a/V_g relationship for analysis of oscillator performance.

For an oscillator under power control but with unchanged drive coupling, it may be expected that the slope of the operating line AA' remains constant and the lines at lower powers are parallel to the original.

Practical measurements of mean anode or gate current may be compared with theoretical assessments obtained by summing the incremental values read off the characteristic curves at points A, B, C, D, E, F, G, F', E', D', C', B' and A' which are at cos 15° intervals. (More frequent intervals increase accuracy but not significantly so.) Thus

$$I_{\text{mean}} = \tfrac{15}{360}[(A+A')+2(B+B')+2(C+C')+\ldots+ +2(F+F')+2G].$$

For class AB2 this complete formula must be used, but for classes B and C the formula simplifies to[4]

$$I_{\text{mean}} = \tfrac{1}{12}[\tfrac{1}{2}A+B+C+D+E+F+G].$$

The same incremental readings of anode and gate current are used to calculate the peak fundamental components of these currents from the formula

$$I_1 = \tfrac{1}{12}[(A-A')+2\cos 15°(B-B')+2\cos 30°(C-C')+ +2\cos 45°(D-D')+2\cos 60°(E-E')+ +2\cos 75°(F-F')].$$

Normally, a transparent film is used on which is printed a graticule comprising a series of parallel lines ranging in length but all divided in the same cosine ratios. This graticule is placed over the characteristic curves and eliminates the need to draw on the graph itself.

If a study begins with a paper design, R_g is chosen to obtain the bias level selected at G from the product $I_g R_g$, where I_g is the mean gate current calculated.

Where analysis of a working oscillator is required, point G is fixed by the d.c. anode voltage and the product of R_g and the I_g reading obtained from the meter. Point A is not as readily fixed. The V_g peak level may be obtained using a suitable peak-reading voltmeter or, if a power controller is effectively clipping the drive voltage, V_g peak $= V_c + \Delta V$ where ΔV is the voltage drop across the clamping diode and its series resistor and is typically about 50 V. The precise value of $V_{a(\text{min})}$ is not readily obtained by measurement but some trial at positioning A will usually yield a condition which gives mean I_a and I_g readings in close agreement with meter readings.

If output power can be measured, for example with a water-cooled calorimeter as load, a further check on the analysis of working conditions can be made since the output power

$$P_0 = \tfrac{1}{2}(V_s - V_{\text{min}}) \cdot I_{a1}$$

where V_s = anode supply voltage

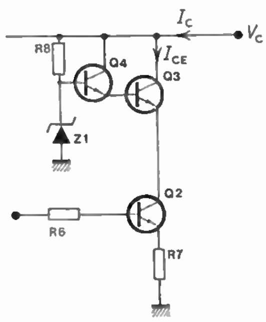
V_{min} = lowest anode voltage reached

I_{a1} = peak fundamental anode current.

In practice about 85% of this power reaches the load. The other 15% is absorbed as heat losses in the oscillator circuit.

Such theoretical assessments combined with practical measurements have shown good correlation and confirmed the action of the gate limiter on the operating conditions.

13 Appendix 2: Control regulator ratings boundary (Figures 6, 8 and 11 refer)



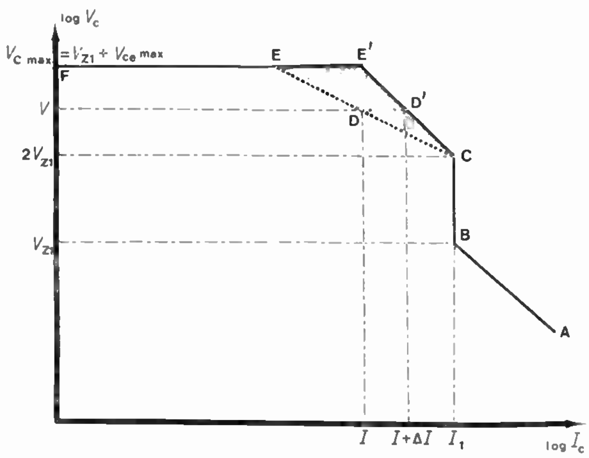


Fig. 11. V_c/I_c relationship for control regulator.

BOUNDARY SECTION AB

High-power transistors usually have a maximum current rating in excess of 1A and the pertinent limit for V_{CE} greater than 30 V is the maximum dissipation. This will be related to heat-sink area and ambient temperature range. At voltage V_c below the Zener diode chain voltage V_{z1}, transistor Q2 takes virtually all of the voltage and AB is its dissipation alone. V_{z1} is less than the V_{CE} maximum rating of Q2.

$$I_1 = P_{Q2}/V_{z1}.$$

BOUNDARY SECTION BC

At voltages V_c above V_{z1}, transistor Q3 takes a share of the voltage and its dissipation limit is reached also at C when it drops an equal voltage V_{z1}, making the total $V_c = 2V_{z1}$ at C where

$$P_c = P_{Q2} + P_{Q3} = 2V_{z1}I_1.$$

To be precise, Q3 handles about 10% less current at its collector than Q2 because Q2 also passes the Q4 collector current. For simplicity in the formulae, Q3 current is assumed to be the same as Q2 current and errs on the side of additional safety.

BOUNDARY SECTION CDE

At voltages $V_c > 2V_{z1}$, transistor Q3 takes the larger share of the voltage and the common current level at the limit must be reduced to keep the dissipation of Q3 within its limit.

At an arbitrary point D, the dissipation of Q3 will be at its maximum rating, $P_Q = I(V - V_{z1})$ and that of transistor Q2 will be IV_{z1} making a total limit of $P = P_Q + IV_{z1}$ (which is also equal to IV as must be the case).

BOUNDARY SECTION CD′E′

At voltages above $2V_{z1}$, the additional current ΔI via the resistor feedback chain R2, R3, becomes more significant and extends the regulator total dissipation boundary to CD′E.

At D:

$$\Delta I = \frac{V}{R2 + R3} \simeq \frac{V}{R2} \text{since } R3 \ll R2$$

and the tota. dissipation $= V(I + \Delta I) = P_Q + V_{z1}I + \dfrac{V^2}{R2}.$

BOUNDARY SECTION E′EF

Eventually transistor Q3 will reach its maximum V_{CE} limit when

$$V_c = V_{CEmax} + V_{z1}.$$

The ratings boundary is then independent of current.

Manuscript first received by the Institution on 30th June 1975, in revised form on 27th January 1976 and in final form on 21st June 1976. (Paper No. 1746/CC266).

UDC 621.372.412

Transformation for modifying the lumped-element equivalent circuit for metal-encapsulated crystals in unbalanced semi-lattice filters

J. K. STEVENSON, B.Sc., Ph.D., A.F.I.M.A., M.Inst.P., C.Eng., M.I.E.E., M.I.E.R.E.*

SUMMARY

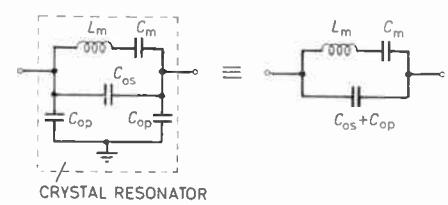
A transformation is given which enables crystals of equal motional inductance to be used in unbalanced semi-lattice filters which specify different inductance values. Allowance is made for stray capacitance between the crystal electrodes and earth; this capacitance is especially significant for metal-encapsulated units.

* *Department of Electrical and Electronic Engineering, Polytechnic of the South Bank, Borough Road, London, SE1 0AA.*

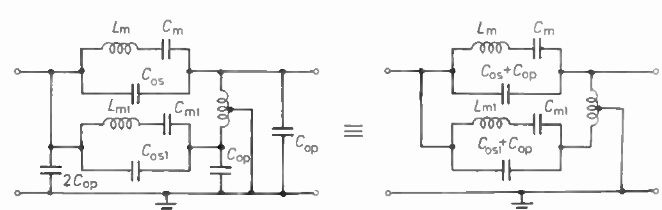
1 Introduction

The lumped-element equivalent circuit for a crystal resonator is shown in Fig. 1(a). C_{0S} denotes the static capacitance and C_{0P} the stray capacitance between each electrode, including the terminal, and earth. Balanced lattice sections are usually realized in unbalanced semi-lattice form (due to Jaumann) as shown in Fig. 1(b); the number of crystals is then halved at the expense of a centre-tapped transformer for each pair of sections.[1,2] In an unbalanced section, C_{0P} can be transferred to across the crystals and this results in an equivalent static capacitance $C_{0S}+C_{0P}$.

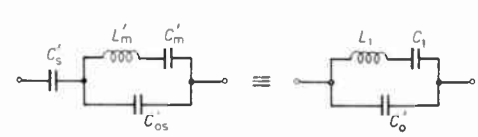
The motional inductance L_m of crystal units which resonate at any specified frequency is restricted to a narrow range of values, often no more than a factor of two; energy-trapping theory (for eliminating spurious responses) sets an upper limit for the electrode mass-loading[3,4] and electrode conductivity sets a lower limit. Crystal filter designs often specify a wide range of values for L_m, so it is frequently necessary to apply a circuit transformation to alter the specified crystal inductances.



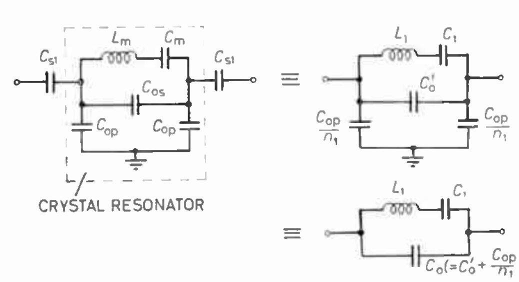
(a) Lumped-element equivalent circuit for a crystal unit.



(b) Transfer of parallel capacitance.



(c) Transformation to increase apparent motional inductance (C_{OP} ignored).



(d) Transformation to increase apparent motional inductance (C_{OP} included).

Fig. 1. Crystals in unbalanced semi-lattice section.

If a series capacitor is introduced to modify the equivalent motional parameters for a crystal resonator, using the well-known transformation in Fig. 1(c),[2] the central parallel capacitor C_{OP} due to the resonator (which is not allowed for in the transformation) distorts the filter response. A new transformation is given here which eliminates this distortion; a symmetrical structure with two series capacitors is used, and transformations are applied to transfer the stray capacitance to the output terminals followed by removal of the series capacitance, as shown in Figs. 1(c) and (d).

2 Transformations

Three cases are considered. The third case, dealt with in Section 2.3, forms the basis of this paper; the initial two cases are included for completeness and to provide formulae for use in the later Section.

The transformations can only be used to reduce the values of crystal inductance specified by the designer. Therefore, if the smallest design value for crystal inductance $(L_1)_{min}$ is less than the smallest permissible crystal inductance $(L_m)_{min}$, all impedances in the design must be increased by a factor of not less than

$$(L_m)_{min}/(L_1)_{min}.$$

2.1 Crystal Design Inductance Satisfactory, $C_{OP} \geqslant 0$

For semi-lattice arms consisting of one or more crystals in parallel, C_{OP} is of no consequence and simply increases the effective crystal static capacitance. Note that for design purposes the effective parallel capacitance is $C_{0S}+C_{OP}$, as seen from Fig. 1(b), and not $C_{0S}+\frac{1}{2}C_{OP}$, the value measured between the electrode terminals of a crystal in an unearthed metal container.

2.2 Crystal Design Inductance Too Large, $C_{OP} = 0$

A transformation for increasing the apparent motional inductance of crystal resonators is given in Fig. 1(c), based on the Norton equivalence in Fig. 2(a).[1] With $n_2=\sqrt{L_1/L'_m}$, the element values are related as follows,[2]†

$$L_1 = \left(1+\frac{C'_{0S}}{C'_S}\right)^2 L'_m$$

$$C_1 = \frac{C'^2_S C'_m}{(C'_{0S}+C'_S)(C'_{0S}+C'_S+C'_m)}$$

$$C'_0 = \frac{C'_{0S}C'_S}{C'_{0S}+C'_S}$$

Hence,
$$L'_m = \frac{L_1}{n_2^2}$$

$$C'_m = \frac{n_2^2 C'_0 C_1}{C'_0-(n_2-1)C_1}$$

$$C'_{0S} = n_2 C'_0$$

$$C'_S = \frac{n_2 C'_0}{n_2-1}.$$

† Note that a prime denotes an intermediate value when the transformation given here is used as part of the more elaborate transformation in Section 2.3. The priming of symbols can therefore be ignored when these formulae are used on their own, i.e. when $C_{OP} = 0$.

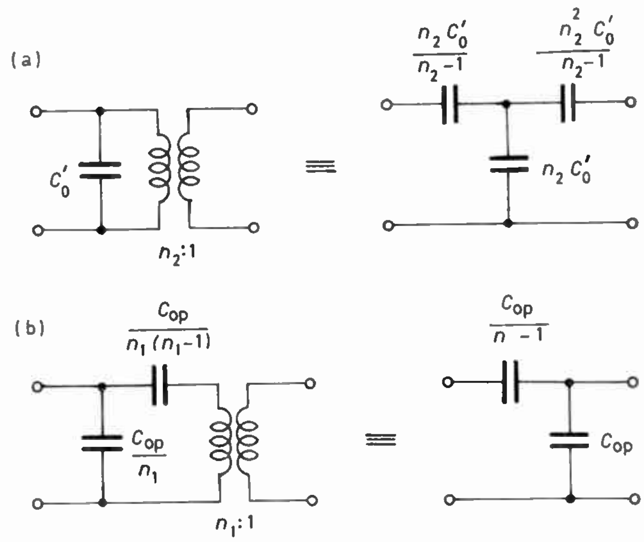


Fig. 2. Norton equivalences.

This transformation is usually suitable for glass-encapsulated crystals, but for crystals in metal enclosures, the central capacitor C_{OP}, in association with the series elements, serves as a potential divider and distorts the filter response.

2.3 Crystal Design Inductance Too Large, $C_{OP} > 0$

To eliminate the distortion resulting from C_{OP}, a series capacitor C_{S1} is connected to each of the crystal terminals, as shown in Fig. 1(d). The Norton equivalence in Fig. 2(b) is applied to transfer the stray capacitors, modified to C_{OP}/n_1 to the output terminals; the crystal parameters L_m, C_m and C_{0S} now become L'_m, C'_m and C'_{0S}. The transformation in Fig. 1(c) then removes the new series capacitors, and provides the values of L_1, C_1 and C'_0 in Fig. 1(d).

From Figs. 1(c), 1(d) and 2, and the formulae in Section 2.2, we obtain the following,

$$L'_m = n_1^2 L_m = \frac{L_1}{n_2^2}$$

$$C'_m = \frac{C_m}{n_1^2} = \frac{n_2^2 C'_0 C_1}{C'_0-(n_2-1)C_1}$$

$$C'_{0S} = \frac{C_{0S}}{n_1^2} = n_2 C'_0$$

$$C'_S = \frac{C_{OP}}{2n_1(n_1-1)} = \frac{n_2 C'_0}{n_2-1}.$$

L_1, C_1, L_m, C_{0S} and C_{OP} are specified by the designer and the remaining circuit elements are given in terms of n_1 and n_2 as follows,

$$n_1 = \frac{2C_{0S}+\sqrt{\frac{L_1}{L_m}}\,C_{OP}}{2C_{0S}+C_{OP}}$$

$$n_2 = \frac{1}{n_1}\sqrt{\frac{L_1}{L_m}}$$

$$C_0 = C_0' + \frac{C_{OP}}{n_1} = \frac{C_{OS}}{n_1^2 n_2} + \frac{C_{OP}}{n_1}$$

$$C_m = \frac{n_1^2 n_2^2 C_{OS} C_1}{C_{OS} - n_1^2 n_2 (n_2 - 1) C_1}$$

$$C_{S1} = \frac{C_{OP}}{n_1 - 1} = \frac{2C_{OS}}{n_1(n_2 - 1)}.$$

The crystal frequency f_m and design frequency f_1 are therefore related by

$$f_m = \left[1 - \frac{n_1^2 n_2 (n_2 - 1) C_1}{C_{OS}}\right]^{\frac{1}{2}} f_1$$

$$\simeq \left[1 - \frac{n_1^2 n_2 (n_2 - 1) C_1}{2C_{OS}}\right] f_1.$$

3 Example

Suppose that the measured parameters of optimally plated 12 MHz crystal units are as follows,

$$L_m = 10 \text{ mH}, \quad C_{OS} = 4{\cdot}5 \text{ pF}, \quad C_{OP} = 0{\cdot}5 \text{ pF}.$$

Let one of these units be used in a filter design which specifies

$$f_1 = 12 \text{ MHz}, \quad L_1 = 100 \text{ mH}.$$

Values for the crystal frequency and the series capacitor C_{S1} are determined using the formulae in Section 2.3, i.e.

$$n_1 = 1{\cdot}1138, \quad n_2 = 2{\cdot}8392$$

leading to

$$f_m = 11{\cdot}984797 \text{ MHz}, \quad C_{S1} = 4{\cdot}39 \text{ pF}, \quad C_0 = 1{\cdot}73 \text{ pF}.$$

4 References

1. Zverev, A. I., 'Handbook of Filter Synthesis' (Wiley, New York, 1967).
2. Temes, G. C. and Mitra, S. K., 'Modern Filter Theory and Design' (Wiley, New York, 1973).
3. Curran, D. R. and Koneval, D. J., 'Energy trapping and the design of single and multi-electrode filter crystals', Proceedings of the 18th Annual Symposium on Frequency Control, Fort Monmouth, 1964, pp. 93–119.
4. Shockley, W., Curran, D. R. and Koneval, D. J., 'Trapped-energy modes in quartz filter crystals', *J. Acoust. Soc. Amer.*, 41, pp. 981–93, 1967.

Manuscript received by the Institution on 30th March 1976. (Short Contribution No. 181/CC287).

Contributors to this issue

John Stevenson (Member 1974) is a Lecturer in the Department of Electrical and Electronic Engineering at the Polytechnic of the South Bank, London. A graduate of the University of Birmingham, he obtained his doctorate following work at Queen Mary College, London. Before taking up his present appointment he was with the General Electric Company, initially at the Hirst Research Centre, Wembley, and subsequently with Marconi Space and Defence Systems. His main research interests are in circuit theory and he has contributed several papers to the Institution's Journal on this subject over the past three years.

Trygve Gytre received his M.Sc. in electronic engineering from the Norwegian Technical University in 1964, and he then lectured in the Department of Telecommunications at the University for one year. In 1966 he joined the Christian Michelsen Institute and worked for the next six years in the field of biomedical instrumentation in the Department of Applied Physics. Since 1973 Mr. Gytre has been at the Institute of Marine Research working for his Doctorate, his research being in sound velocity and in applying oceanographic instrumentation, the latter in close cooperation with Chr. Michelsen's Institute.

UDC 551.465: 534-8

The use of a high sensitivity ultrasonic current meter in an oceanographic data acquisition system

TRYGVE GYTRE, M.Sc., M.N.I.F.*

Based on a paper presented at the IERE Conference on Instrumentation in Oceanography in Bangor from 23rd to 25th September 1975

SUMMARY

The paper describes the design of a high performance two-axis ultrasonic current meter and how this sensor has been designed into a general oceanographic data acquisition system with additional sensors for pressure, temperature, salinity, sound velocity and current direction.

* *Institute of Marine Research, Bergen, Norway*

1 Introduction

The general aim of physical oceanography is to describe the physical properties and movements of the seawater and its interactions with the surroundings.

Within this science, which has both purely scientific and practical applications, a description of the water movements or the currents is very central for the following main reasons:

1. Currents in the sea exert physical forces on human products which work in or close to the sea.

2. Currents act as a carrier for thermal and kinetic energy (climatic aspects), for dissolved nutrients (biological aspects), and for human wastes (pollution aspects).

In order to understand the nature of the currents, i.e. whether the current is a part of a wave motion which develops forces but no net transportation, or a part of a turbulence which mixes it with surrounding water, a quantitative description of the current magnitude and direction is necessary. Traditionally current is measured by a mechanical sensor which obtains a rotation or a drag from the moving water in proportion to the water velocity. Mechanical sensors, however, become unreliable when the water velocities approach a few centimetres per second. Their bandwidths are also limited to about 0·25 Hz which make them unsuited for fast turbulence studies.

Recent developments in electronics and transducer technology, however, have now made it possible to measure fluid velocities by exploiting the interaction effects between a moving fluid and either an electromagnetic field or acoustic waves. A number of prototype instruments designed around non-moving sensors have been built, but until now no particular instrument has emerged significantly over its competitors in quality. At present electromagnetic current meters are in some general use. Laser Doppler current meters have been developed, but they are still only suited for very special purposes where an abundance of power is available. A few ultrasonic current meters have also been developed. The latter use either a Doppler principle or the travel time difference principle.

A group in Bergen centered around Chr. Michelsen's Institute and the Marine Research Institute has for some time worked on an ultrasonic current meter based on the travel time difference principle. At present this group has succeeded in developing the electronic circuits and mechanical components necessary to measure simultaneously several components of current down to 1 mm/s. In addition a general data acquisition system for 16 analogue sensors and any practical number of digital sensors has been designed. The data are recorded on a digital cassette tape recorder, or on an audio cassette tape recorder.

2 Current Meter Design Principle

The travel time difference principle for measuring currents is illustrated in Fig. 1. By exciting two piezoelectric transducers with sharp voltage steps of 200–300 V at a repetition rate of, say, 30 excitations per second

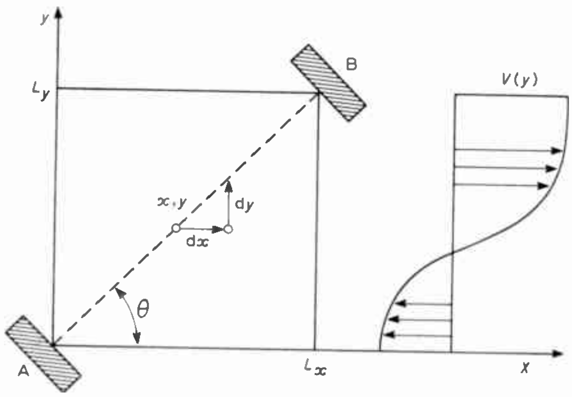


Fig. 1. Travel time difference principle.

simultaneously, each of the transducers generate bursts of high frequency ultrasound—typically 4 MHz—at the same repetition rate.

Figure 1 shows two such transducers which are pointed towards each other in a one-dimensional flow field $v(y)$. The sound path between the two crystals is at an angle θ to the x-axis which coincides with the flow direction.

The travel time for the leading edge from A–B is given by

$$T_{AB} = \int_A^B \frac{dl}{v(y)\cos\theta + c} \simeq \int_A^B \frac{dl}{c} - \int_A^B \frac{v(y)\cos\theta}{c^2}\,dl \quad (1)$$

provided $v(y) \ll c$ where c denotes the speed of sound in the fluid. Another wave train is simultaneously generated at B. The transit time from B to A is

$$T_{BA} = -\int_B^A \frac{dl}{v(y)\cos(\theta+\pi)+c}$$

$$\simeq \int_B^A \frac{dl}{c} + \int_B^A \frac{v(y)\cos\theta}{c^2}\,dl. \quad (2)$$

The transit time difference is then:

$$\Delta t = T_{BA} - T_{AB} = 2\int_A^B \frac{v(y)\cos\theta}{c^2}\,dl. \quad (3)$$

Defining the mean velocity along the signal path as

$$\bar{v} = \frac{1}{L}\int_A^B v(y)\,dl \quad (4)$$

the transit time difference equals:

$$\Delta t = \frac{2L\bar{v}}{c^2}\cos\theta. \quad (5)$$

According to (5) the travel time difference is a function of the fluid velocity, the sound velocity and the projected length of the sound path.

3 Designing for Best Possible Sensor Performance

When mounted the transducers constitute a current sensor. This sensor should be designed to give optimum hydrodynamic performance at all velocities and current directions wanted.

In short we require that:

1. The sensor responds selectively to the component of current along each transducer axis.

2. The response of each transducer pair is maximum when aligned with the current and falls off as the cosine of the angle between the transducer axis and the flow direction when the sensor is rotated.

3. The sensor does not disturb the water where the measurement takes place.

4. The sensor should be able to measure accurately both static and dynamic currents from 1 mm/s to approximately 2·5 m/s.

Assuming a value of L of, say, 15 cm and a sound velocity of 1500 m/s, the demands listed above mean that Δt must be determined by the electronic part of the instrument to a precision of 10^{-1} seconds.

To obtain a good signal-to-noise ratio when detecting the signals, the two transducers must be mounted along their acoustical axis at an internal distance that preferably provides a maximum received signal intensity. For a circular transducer the intensity along the axis is given by:

$$I = I_0 \sin^2 \frac{k}{2}\left\{\left(Z^2 + \frac{D^2}{4}\right)^{\frac{1}{2}} - Z\right\} \quad (6)$$

where k = wave number,

Z = distance along axis,

D = transducer diameter,

and the maximum intensity occurs at

$$Z = \frac{D}{4\lambda} = \frac{S}{n\lambda} \quad (7)$$

where S = transducer surface area.

Figure 2 shows a plot of the function.

The sensor design thus becomes a compromise between the need to obtain a good signal-to-noise ratio, the need to keep the transducers as small as possible to avoid disturbing the flow, and their internal distance long to obtain a high Δt-value. The typical compromise is to chose a probe distance approximately equal to the pressure container diameter which must be between 12 and 15 cm to make room for a tape recorder. Using piezoelectric crystals of 10 mm diameter, the received signal is approximately one-third of the maximum

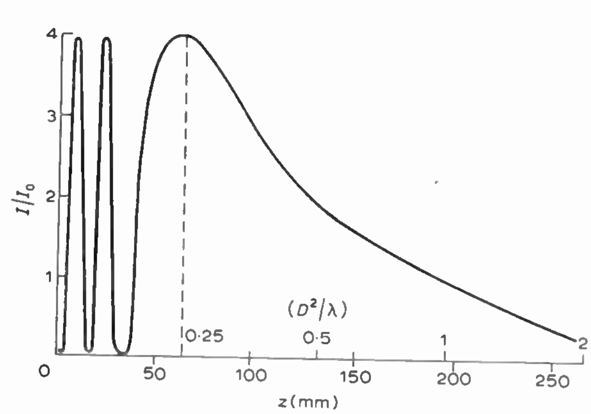
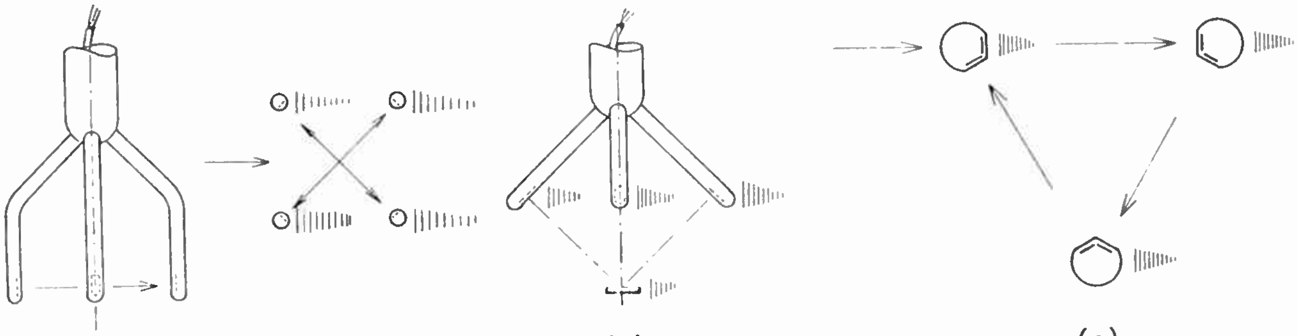


Fig. 2. Relative acoustic intensity as a function of distance between transmitter and receiver.



By orientating the acoustic path away from the current direction.

By using a reflector.

By using 3 paths and neglecting signal from the one that is disturbed.

Fig. 3. Principles for reducing effect of circulation and separation behind the transducers.

obtainable—which is regarded tolerable. For protection against moisture and pressure the crystals are mounted in acoustical contact with the inside of a metal disk facing the current. When excited the acoustic signals penetrate the protective disk and propagate through the water.

4 Flow Conditions Around the Probes

To obtain a true measure for the flow field, the transducers should preferably have zero dimensions. Some practical sensor designs are shown in Fig. 3. All the designs employ piezoelectric crystals of 10 mm in diameter mounted in 15 mm thick tubes. The distance between two crystals in a transducer pair is typically 150 mm. When using the design shown on Fig. 3(a), the wake behind the leading transducer may modulate the velocity field and thus contribute to an error in the final result. A complete description of the conditions around the probes at laminar, transition and turbulent flow conditions is very complicated. An approximate description has been approached by using potential flow theory. The effect of flow separation has been approximated by adding circulation around the cylinders in which the crystals were mounted. By integrating the average vector component along the ultrasonic signal path and assuming that the output signal from the electronic part of the instrument is proportional to the length of the signal path, the expression for the instrument output will be:

$$V = kv_0(s_1 + s_2\,|\sin\theta|)\cos\theta \qquad (8)$$

where V = signal from the instrument,

v_0 = undisturbed stream velocity,

s_1 and s_2 are constants,

k = conversion factor between analogue output signal from instrument and fluid velocity,

θ = angle between velocity vector and ultrasonic beam direction.

Using a computer to do curve fitting to an observed behaviour of a current meter with probe arrangement as in Fig. 2, it was found that the current signal follows the true cosine response for values of θ greater than approximately $0{\cdot}2\pi$, or 25°. Smaller values of θ makes the instrument show too small values for the current.

To obtain a correct measure for the current, the ultrasonic beam hence should cross the vector at an angle of more than $\sim 25°$. This can be obtained in three ways:

1. By the use of a vane to direct the acoustic signal in a wanted angle referred to the current direction (Fig. 3(a)).
2. By deviating the ultrasonic beam between the transducers into a 'V'-shape using a reflector (Fig. 3(b)).
3. By using three signal paths in a triangular mode and neglecting the signals from the most disturbed signal path (Fig. 3(c)).

Till now configuration 2 has been favoured. Figure 4 shows the design in detail. Ultrasonic signals are simultaneously emitted from four transducers at an inclined angle of 45° with the horizontal plane. The transducer tips have been shaped to make the inevitably increased local currents around them cross the ultrasonic signals at an angle of 90°.

Approach No. 3 demands a less sophisticated sensor design but a more elaborate electronic solution. However, as electronic components can do more and more for less and less money, this represents an obvious evolutionary trend.

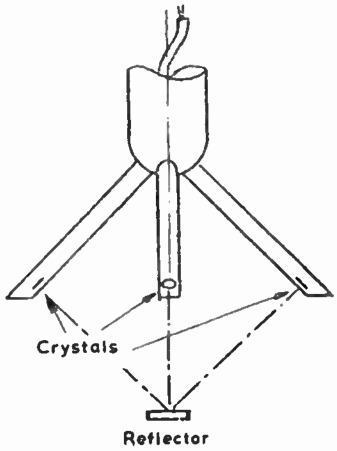


Fig. 4. Two-axis current transducer with reflector.

5 Electronic Design

To measure a velocity of 1 mm/s and be sure that the measurement is real, the current meter noise level should be less than that corresponding to 0·5 mm/s. Practically then a time difference of less than 100 ps must be measured. Such small time differences are very easily modified as much by changes in the electronic components themselves as by small physical changes in the water. To be sure that it will measure correctly, the circuit must therefore, in addition to the time difference measurements, also measure and compensate for the effect of changes in the water sound velocity and for drift in the electronic circuits.

Figure 5 shows a block diagram of how this is achieved. The circuit is controlled by an oscillator whose frequency is divided in a dividing chain. Each time the second from last stage in the chain goes high, the transducers are excited. The last stage is two or more decoded outputs from a counter that in succession enables the receiver and sampling circuits that correspond to one particular current axis.

As the counting chain is all at zero during crystal excitation, the counter status when the signals are received at their opposite transducers is a measure for the travel time and hence a measure for the sound velocity. The four least significant bits in the chain are copied in a latch and used to compensate for changes in the sound velocity according eqn. (5) by modifying the total signal gain.

In order to overcome the slow response of digital logic circuits, the minute differences in travel time are detected by letting the received signal from one crystal start a high-speed ramp generator and the signal from the opposite crystal stop it via a delay-vibrator M3. The ramp amplitude is amplified according to the measured sound velocity and sampled in sequence. Between the samplings a simulated zero signal is introduced to the circuit. If zero drift occurs, a compensation signal is added to counteract it. The correct sequence of operation is secured by a sequence control unit which gets its information from the dividing chain.

6 Results

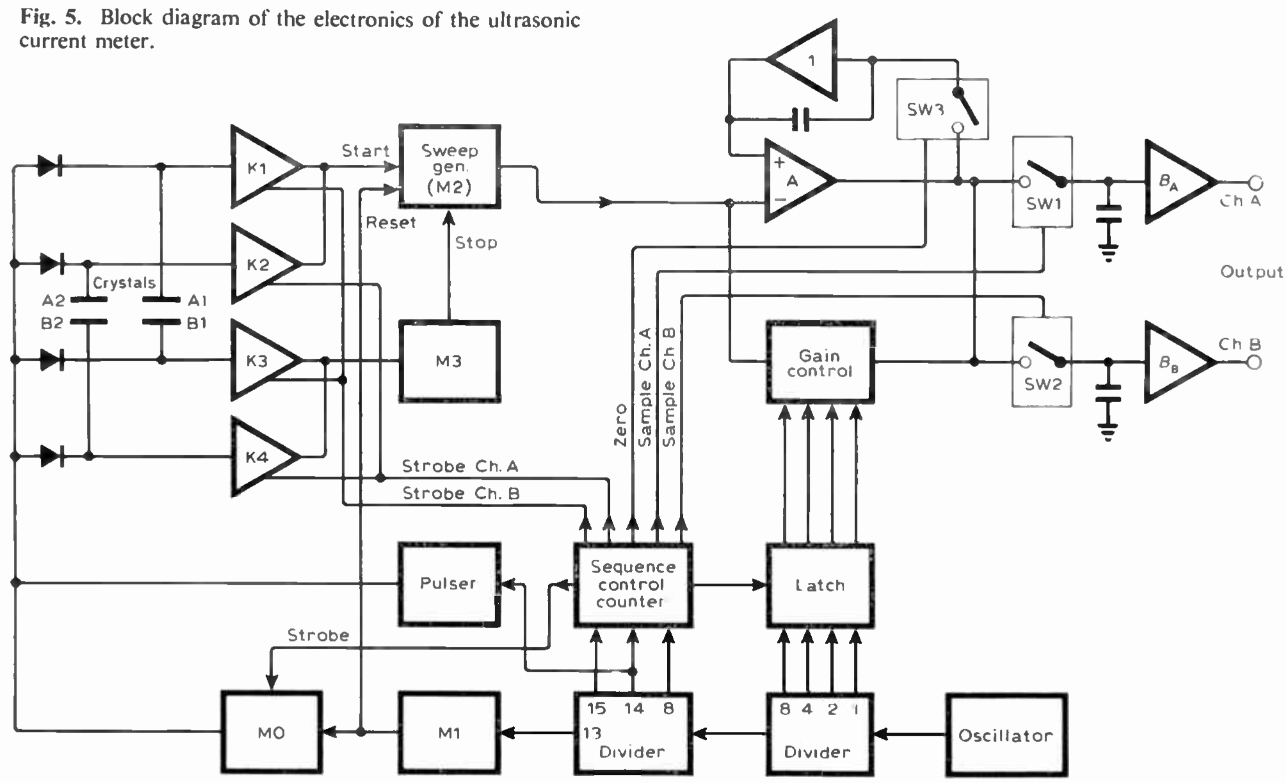
In combination with the two-axis reflecting sensor the circuit described has proved to measure currents very precisely. The noise level is $\sim 0{\cdot}5$ mm/s. Long-term stability over several weeks as measured in the laboratory is better than a few mm/s, and the dynamic range is $\sim \pm 2{\cdot}5$ m/s.

The dynamic response of the instrument is a function of both the spatial integration described in eqn. (1), the electric integration performed in the output sample-and-hold circuits, and the instrument sampling rate. The maximum bandwidth according to the sampling theorem is equal to the sampling frequency divided by two. At present the typical sampling rate is between 20 and 30 Hz per channel resulting in a typical 3 dB bandwidth of about 5 Hz.

Calibration experiments with a sensor designed according to Fig. 3(b) have been performed in a towing tank at the Technical University of Trondheim. Figure 6(a) and (b) presents the results from one calibration experiment showing the response from both changes in velocity and values of θ.

The design described can in principle be extended to an unlimited number of axes, making, for example,

Fig. 5. Block diagram of the electronics of the ultrasonic current meter.



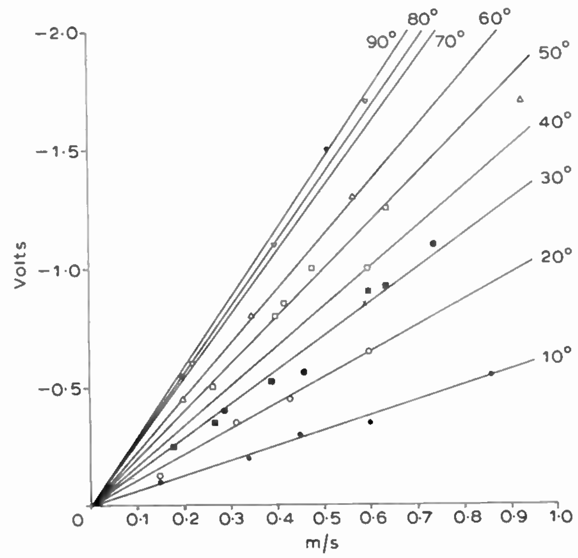


Fig. 6. Calibration diagram for reflector type transducer.

measurement of shear currents in several directions possible. Till now, however, only versions with up to three axes have been made.

7 Direction

No current is adequately described until its direction has been given.

To measure the instrument orientation a simple saturated core compass has been made, and Fig. 7 shows its principle. A toroidal core of high permeability iron is excited at a frequency of about 1 kHz, and at a given phase angle the core saturates. If the core is homogeneously magnetized, a pick-up coil around the entire magnetic circuit will detect no signal. However, when the Earth's magnetic field adds to the excited flux, the field distribution in the core becomes unsymmetrical and a second harmonic signal component with amplitude proportional to the external field intensity develops.

A phase-sensitive demodulator extracts the amplitude and sign of the second harmonic signal and feeds it back to the detector coil to generate a magnetic field which is equally strong and has a direction opposite to the incoming component of the magnetic field. Finally, the sensor output is limited to a convenient maximum value of, say, ± 5 V p-p.

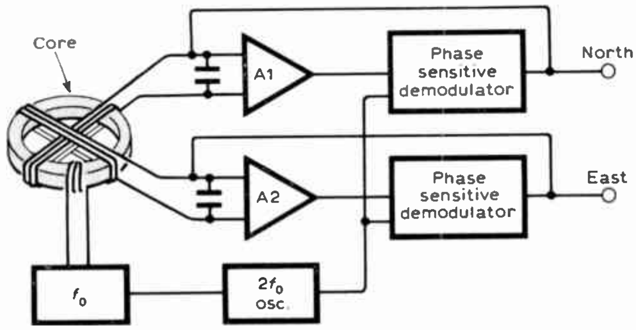


Fig. 7. Principle of fluxgate type compass.

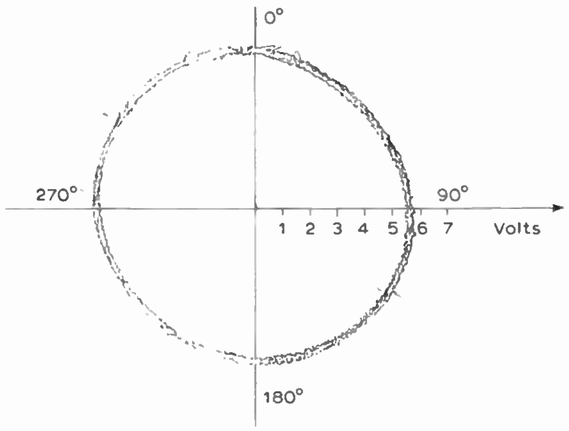


Fig. 8. Compass response on x–y recorder. The compass has been rotated 360° three successive times at a tilt angle of 0·5 and 10° respectively.

Two orthogonally wound pick-up windings detect a signal which is proportional to the sine and the cosine of the angle of rotation respectively. These signals are denoted by NORTH and EAST, and the angle of rotation may be obtained by computing arctan NORTH/EAST.

Alternatively, the signals from the current meter may be combined with those from the compass according to the equations:

$$\begin{aligned} V_N &= V_x \cos\theta - V_y \sin\theta \\ V_E &= V_x \sin\theta + V_y \cos\theta. \end{aligned} \tag{9}$$

The compass signals are thus suitable for analogue multiplication which can be done in the instrument to make possible vectorial integration and thus save data recording space.

As the magnetic field is generally inclined with reference to the horizontal plane, the compass sensor must always be kept horizontal. The sensor is therefore suspended from an universal joint and damped in oil.

Provided the horizontal position is known, the compass has a resolution limit of 0·1 degree. Figure 8 shows test results from a compass where the NORTH and EAST components have been fed to an $x-y$ recorder and the compass rotated 360° at different tilt angles. Due to unsymmetrical winding this output is not a perfect circle.

8 Other Sensors

To describe the oceanographic environment, information on a number of additional variables is necessary. At present work has been done to measure temperature, pressure, salinity, sound velocity and oxygen.

8.1 Temperature

Temperature is measured by a combination of a platinum thermometer and a thermistor. The platinum thermometer is protected in a metallic housing and measures with long-term stability, but with a time-constant of more than 0·6 s. To improve the response time, an unprotected thermistor with response time of about 0·1 s is added. Typical precision in temperature measurements is $\sim$0·01 deg C in the range -2 to $+18$°C.

8.2 Pressure

A standard strain-gauge-type pressure transducer is used for depth indication. At present integrated circuit transducers like National Semiconductor LX1600 have been used to some extent. This transducer has a built-in amplifier which makes succeeding signal shaping easy.

8.3 Salinity

The standard ways to measure salinity make use of inductive coupling or measurement of the water conductivity. Both methods are sensitive to fouling of the transducer facing, and high-precision long-term monitoring of salinity is therefore not possible.

To obtain a salinity sensor with long-term stability, experiments are now performed with a travel-time-difference sensor, in which two ultrasonic signals are propagated over the same fixed distance. One of the signals travels in an enclosed tube of standard 35‰ sea water, while the other travels in the unknown sea water outside. The difference in arrival time is measured in the same kind of circuit as used in the current meter. The difference in travel time ΔT_t thus becomes a measure for the difference in inverse sound velocity ($l/c_u - l/c_{35}$) where c_u is the sound velocity in the surrounding water and c_{35} is the sound velocity in the standard water.

As described by, for example, Wilson,† sound velocity is a function of temperature, pressure and salinity. Thus if temperature and pressure are kept equal, the difference in travel time for the two paths becomes a function of salinity only. At present the greatest drawback for this sensor is the long temperature equilibrium time constant ($\sim$10 s). To speed up response time a platinum thermometer along both paths measures both the absolute and the differential temperature.

Provided there is perfect temperature and pressure equilibrium this sensor can resolve differences in sound velocity down to $\sim$1 mm/s which correspond to a change in salinity of less than 0·001‰. With a practical error in the determination of the temperature difference between 0·01 and 0·001 deg C a precision in salinity determination around 0·01‰ will be readily obtainable.

As fouling of the transducer surface will not affect the transmission of sound, this sensor should be well suited for long-term purposes. The sensor is now (April 1976) operating in a tank of sea water with some 20 fish (cod) which are fed once a day in order to study how biological fouling affects the performance. So far the output has proved stable.

Figure 9 shows a first calibration for the salinity range 32–34‰. A one-year measurement of the salinity outside Svalbard with this sensor is planned in 1976–77.

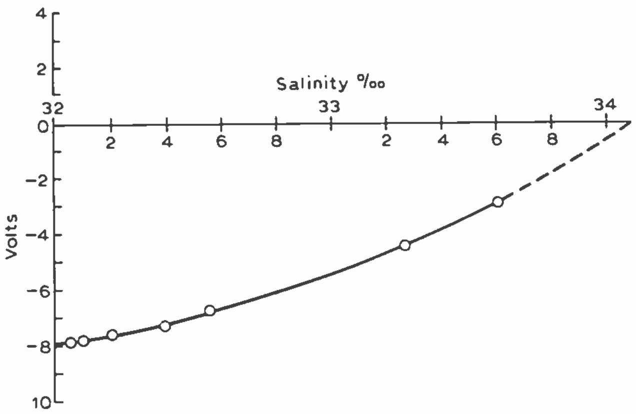


Fig. 9. Calibration for the acoustic salinity sensor.

8.4 Sound Velocity

As sound velocity is a function of density which again is a function of temperature, salinity and pressure, separate measurements of sound velocity, temperature and pressure may provide input data for the calculation of the sea water density.

In the primary stages of the current meter electronics the travel times T_1 and T_2 are readily available. This information is extracted and converted to an analogue signal. In order to obtain a high sensitivity, the travel time is divided into a fixed part T_F and a variable part T_V, where T_V is typically $0{\cdot}1T_F$. T_F is then scaled to give analogue signals between -10 and $+10$ V and c is computed as $l/(T_F + T_V)$, l being the acoustic path length.

As acoustical pulses are emitted at least 20 times per second, the response time for sound velocity is extremely small which makes this sensor ideal for profiling.

8.5 Oxygen

The oxygen sensor has not yet been developed to a useful level of performance. The planned development effort is based on the use of an Au–Zn polarographic electrode. This electrode has an output signal proportional to the stirring rate and the water oxygen saturation. The development aims at obtaining a controlled stirring rate as well as self-cleaning by periodically exposing the electrochemical part of the sensor to high-intensity ultrasonic shock waves.

9 Recording of the Data

Recording can be done digitally or by frequency modulation. For digital recording we have selected a digital tape recorder made by Sea Data Inc., USA. It can record approximately 10^7 bits at a rate of 400 bits/second on a standard C-60 cassette. To do this it requires the data to be available on a serial-out shift register.

The analogue data fed in are connected to a 16-channel multiplexer followed by a 12-bit c-m.o.s. analogue-to-digital converter which again is followed by a 12-bit shift register. The multiplexer is addressed by a counter.

On a record command from a timer, the multiplexer addresses a pre-programmed number of the channels to the a/d converter. The data from the a/d converter are quickly shifted into the shift register and then clocked into the tape recorder by slow 400 Hz shift pulses from the tape recorder.

In the intervals between each 12th and 13th shift pulse, a new analogue channel is converted and the shift

† Wilson, W. D., 'The speed of sound in sea water as a function of temperature, pressure and salinity', *J. Acoust. Soc. Am.*, 32, No. 6, pp. 641–5, June 1960.

register is accordingly refilled. Signals from any digital sensors are simply inserted in a shift register in front of the previous register.

When continuous recording is preferred, a frequency-shift-modulating principle is used. The data stream from the shift register that succeeds the a/d converter is impressed on a frequency shift modulator which gives one oscillation of frequency 1 kHz when a logical one is on the line and two oscillations of frequency 2 kHz when a logical zero is present. The signals are then simply recorded on a standard miniature audio tape recorder.

An identical modulation system is used when the data are to be transmitted along a single cable to the surface. For the demodulation of these signals a simple demodulator providing both 16 analogue channels and digital output has been designed.

10 Timer

To measure representative currents, the current values must be burst sampled at a repetition rate of at least twice the expected wave period during each measurements period. To obtain this a specially-designed timer has been made, in which a crystal-controlled oscillator is frequently divided down to a selectable period time of $T_1 = 1$ to 256 min with a resolution of 1 minute. At a repetition rate of T_1 the electronic circuits are connected to the battery. Then record command pulses at a repetition rate of T_2 seconds are generated. Each record command causes the multiplexer to scan over all the channels that are connected to a sensor, convert the data to digital information and clock them into the tape recorder. After a predetermined number of 'record commands', the power is automatically disconnected from the electronic circuits until the next period of measurements is to start.

11 Results

During the period from 1970–76 approximately 25 current meters with increasing degree of sophistication and performance have been made. A number of the instruments have been in use for years and have proved to work reliably without maintenance.

The most simple instruments contain one two-axis circuit board and one d.c.–d.c. converter board with outputs of two analogue current signals in the range ± 10 V.

Figure 10 shows a self-contained recording instrument for profiling purposes. This instrument measures two axes of current, two axes of compass, sound velocity, pressure, slow and fast temperature and records the information on an audio tape recorder.

12 Discussion

When deciding to measure currents by means of ultrasonic signals, a wide variety of principles for signal generation and signal detection is possible. The travel-time-difference principle used in this instrument has been chosen for its simplicity and versatility in extending the number of axes, and its very wide dynamic range from 10^{-3}–2×10^{0} m/s. Due to the high voltage signals needed for excitation and the fast response needed for

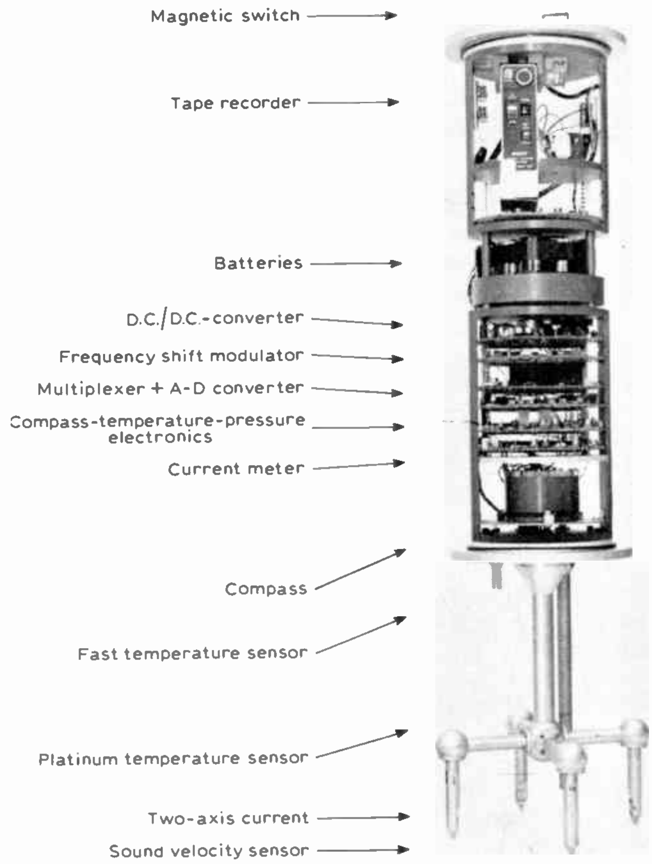


Fig. 10. Self-contained, recording, data collecting instrument with sensors for current, compass, pressure, temperature and sound velocity.

the detection of received signals the power consumption is close to 1 W which is more than an instrument working on the continuous wave principle would need.

However, as batteries are becoming increasingly efficient, the power consumption is not a severe limiting factor. Already now the data capacity in recording versions is limited by the tape cassette capacity rather than by available energy.

The next major re-design of the instrument will involve extended use of c-m.o.s. logic and low-power-consumption operation amplifiers which is expected to significantly reduce the power consumption.

13 Conclusion

Current measurements are at present almost exclusively made by means of rotor type sensors. The sophisticated, low power electronic components which are now becoming increasingly available make it possible to design current meters which can compete with rotor instruments in volume, power consumption and price and by far outperform the rotor instruments in measurement quality. The travel time difference type acoustic current meter will probably play an important role in future current meter design.

Manuscript first received by the Institution on 30th April 1975 and in revised form on 27th May 1976 (Paper No. 1747/AMMS 75.)

Putting Technology to Work

SIR IEUAN MADDOCK, C.B., O.B.E., D.Sc., F.R.S. (*Past President*)*

In an address as Chairman of Committee X at the British Association Meeting in Lancaster on 3rd September 1976, Sir Ieuan Maddock examined the frequently heard comment 'the UK is first-class at inventing things but bad at exploiting these inventions'. He discussed the factors that influence success in 'putting technology to work', and concluded by putting forward eight measures for deriving greater benefit from the national investment in R & D.

Inventive Britain

How inventive and creative are we? This is a difficult question to answer in any quantitative way. Certainly the UK performance in fundamental research has been good ever since the 17th century and particularly good since the early 19th century. One very crude index of the high performance in the fundamental sciences can be derived by comparing the number of Nobel prizes in science-related subjects that have been won by the leading countries of the world. Table 1 shows that relative to its population Great Britain has been an outstanding performer, significantly more than France whilst Japan is only barely significant. Alas, there are no Nobel prizes for engineering, food growing and processing, transport technology or communications, and if there were the balance of the prize winners could be different.

Another way of assessing the inventiveness (or more correctly the 'innovativeness') of countries is to examine their success in introducing commercially viable inventions or discoveries. In recent years there has been no systematic study of this measure of technological success but some years ago a study was undertaken by the OECD which ultimately appeared as a paper by Keith Pavitt (now at the Science Policy Research Unit at the University of Sussex).[1] He examined the success rate of the main industrial countries by six different methods.

(a) Location of 110 important innovations since 1945.

(b) Receipts for patents, licences and knowhow from 6 major countries (1963–64).

(c) Origin of technology imported by Japan (1960–64). This is an ingenious choice and implies that Japan has been adept at selecting good foreign sources to meet her national technological needs.

(d) Patents taken out in other countries.

(e) Export performance in research intensive industries (1963–65).

(f) Export performance in research intensive product groups.

The relevant data are given in the form of histograms in Fig. 1. Pavitt points out that there should be a population correction applied on the basis that a larger population means more inventors, more experimenters and a bigger innovation market. On the assumption that this correction is linear (it may not be) the figures for UK, West Germany and France need not be changed (almost equal populations) but the Japanese figures should be halved and the US figures reduced by a factor of four. On this basis the performance of the UK overall seems not very much different from that of the USA or of Germany (the latter scores well in charts (d), (e) and (f) but significantly better than either France or Japan. There is no overwhelming evidence here that the UK is exceptionally good or bad in producing innovations except in relation to France or Japan but what follows may be significant in relation to these two countries.

Research

For over a century the UK has had a deep preoccupation with research and development and has spent a bigger proportion of its Gross Domestic Product (GDP) on R&D than any other nation of comparable size (since 1945 in aggregate nearly twice as much as Germany and France and almost five times as much as Japan). Currently the UK spends 2·3% of its GDP on R&D and to emphasize the magnitude of this figure it is more than 50% greater than the entire net output of the Coal Industry or directly comparable to the entire net output of the Agriculture Industry. R&D therefore ranks as one of the nation's major 'industries'.

An investment of this magnitude can only be justified if there is a corresponding benefit to the community at large (as opposed to the scientists themselves) in terms of security, industrial performance, social wellbeing etc. But since the ability to maintain security or to afford social benefit services depends ultimately on our industrial strength it is here that the benefits must be seen to fall. Yet over the past two decades those very sectors of British industry that should have benefited from the heavy R&D investment—the engineering-based industries—have become progressively less competitive.

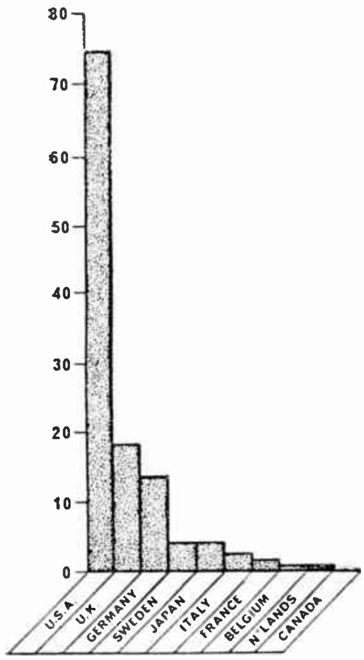
I have discussed this dilemma in some detail elsewhere (the 7th Royal Society Technology Lecture[2]) so I will not repeat the bulk of that material here. I will however highlight a few of the facts in order to establish a basis for what follows.

One of the common illusions is that process of discovering or inventing is a sufficient act in itself and that success will automatically follow. We point with pride to the fact that Rutherford and his collaborators (who were very multinational in origins) laid the foundations of nuclear power. But important though their contribution was, the nuclear

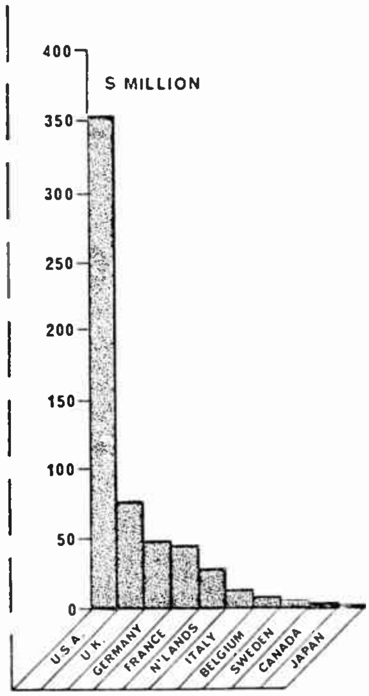
Table 1. Nobel Prizes 1901–1974

	Physics	Chemistry	Physiology and Medicine	Aggregate
	(Total 101)	(86)	(113)	(300)
U.S.A.	32	20	42	94
Great Britain	19	19	17	55
Germany	14	23	10	47
France	9	6	6	21
U.S.S.R	6	1	2	9
Japan	3	—	—	3

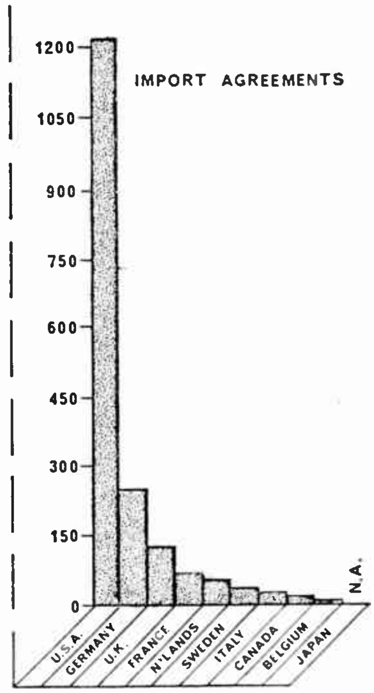
Chief Scientist, Department of Industry, Abell House, John Islip Street, London SW1P 4LN.



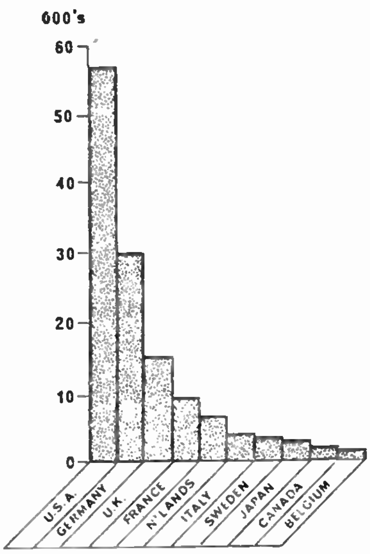
(a) Location of 110 innovations since 1945.



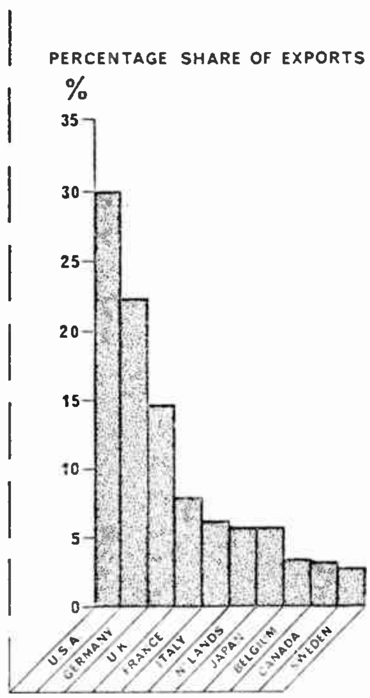
(b) Receipts for patents, licences and know-how from 6 major countries 1963–1964.



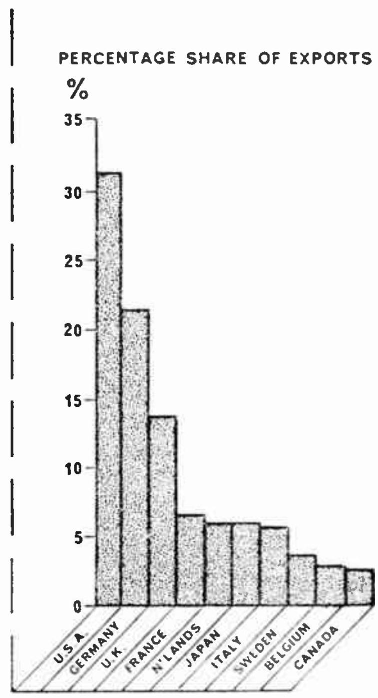
(c) Origin of technology imported by Japan 1960–1964.



(d) Patents taken out in other countries 1963.



(e) Exports performance in research-intensive industries 1963–1965.



(f) Exports performance in research-intensive product groups 1963–1965.

Fig. 1. Measures of technological success.[1]

reactors derive almost wholly from the hundreds of millions of man-hours of research, of development, of painstaking prototype testing by a veritable army of gifted scientists and engineers. Whereas the early work of the pioneers opened certain doors and established some important sign posts, ultimate success called for the creation of a vast new branch of metallurgy, of radiation measurement, of materials preparation and fabrication, of heat transfer, of control theory and a new dimension to mechanical engineering. Throughout there was the need to establish quality control, safety measures and rigorous work disciplines on a scale that had not previously been experienced anywhere. Very little of this work will be recalled in history, none of it will win Nobel prizes yet it is the essential stuff on which nuclear energy is based.

Likewise we can point with pride to the fact that radar emerged primarily from the fundamental work of the then Radio Research Station of the DSIR and from the enterprise of Watson-Watt and his collaborators. Credit-worthy (and timely) though this was, it is wrong to overlook the fact that ultimate success came from the endeavours of a multitude of researchers, engineers, designers, quality control experts etc., who made possible the high-frequency oscillators, the wave-guides and antennae, the pulsed modulators, the wide-band receivers, the c.r.t. displays and a greatly improved understanding of propagation phenomena and information theory.

I will not labour this point further but just emphasize that it is the long hard slog of engineering that leads to final success rather than the initial flash of inspiration.

The UK Economy

I must spend a few moments to make some points about the UK economy in general and the engineering industries in particular. First let us look at the composition of the Gross Domestic Product and in the interest of simplicity this is shown as a composite pie diagram for the years 1950, 1960 and 1970 in Fig. 2. The contents of this chart will probably be a surprise to many since we all carry false impressions based on our own limited personal observations. What is immediately apparent is that there are substantial sectors of economy such as 'others' (nearly 30% of the GDP), distributive trades (a further 10%) and construction (6%) that are only marginally 'technological' in character. Conversely, just because of their size (approximately a half of the GDP) these can make a major contribution to improved economic performance just through the introduction of modest but effective techniques. Examples of these are the influence of earth moving machinery, tower-cranes and ready-mix concrete in the construction industry or of refrigeration and automated packaging in the distributive trades.

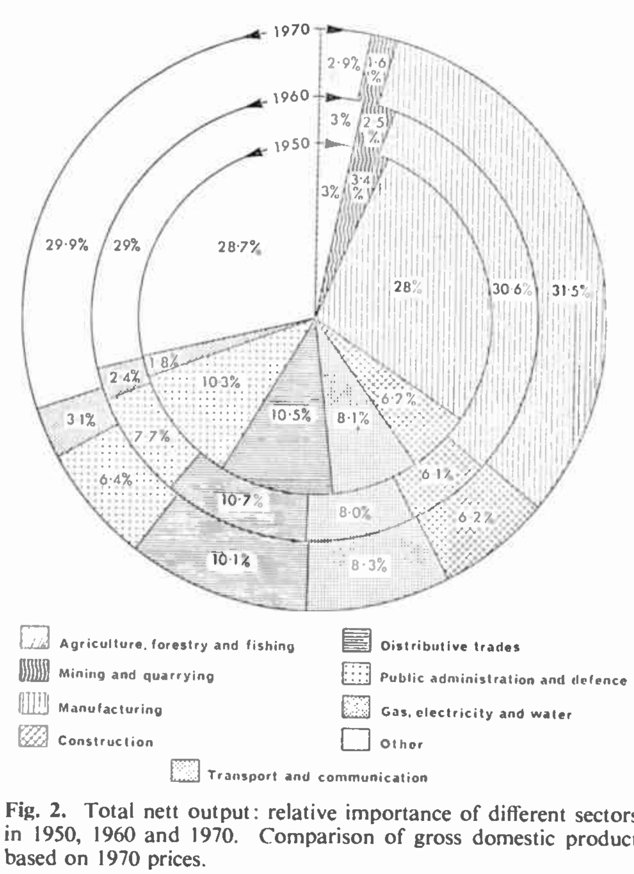


Fig. 2. Total nett output: relative importance of different sectors in 1950, 1960 and 1970. Comparison of gross domestic product based on 1970 prices.

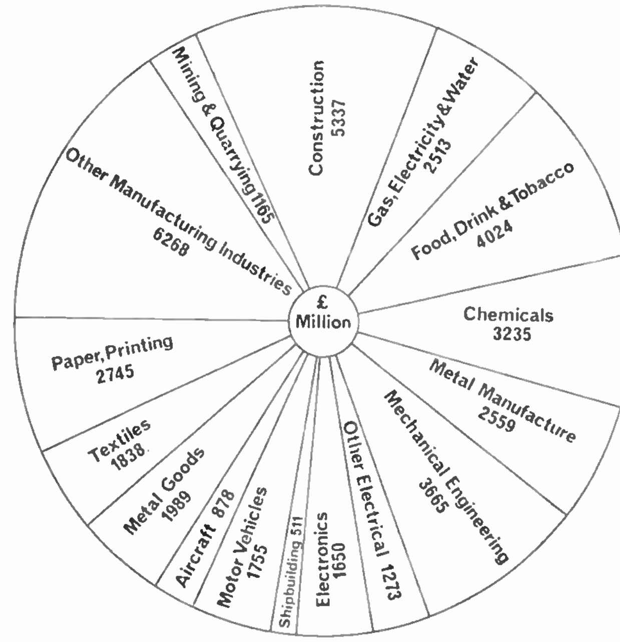


Fig. 3. UK nett output 1974. Total = £41405M
Source: Census of Production 1975.

Turning to the other half of the GDP covering such activities as agriculture, mining, manufacturing and the utilities the technological content is much higher and in some cases is dominant, Even here however, the correlation between the level of the technology and its ultimate benefit to the economy is very variable. Thus a modest technology, e.g. farm tractors and harvesting machinery, can have a profound effect on a significant sector whilst a very advanced technology which is very specific to a narrow sector (e.g. superconductivity), can end up by having no discernible effect. The important quality therefore is the *relevance* of the technology and not its brilliance.

Figure 2 is too coarse to indicate the weight of the different sectors that are particularly relevant to technology and Fig. 3 gives a more detailed breakdown. The significance of the construction industry has already been mentioned but this chart also points out the importance of such sectors as 'other manufacturing industries'*, food, drink and tobacco, mechanical engineering, chemicals, paper and printing, textiles and motor vehicle manufacture. By comparison aircraft and shipbuilding occupy only a modest fraction of the total cake. If technology is to produce a major benefit to the economy it must exert its influence on the very large sectors. There is an old saying which says 'a penny

* Other Manufacturing includes rubber, floor coverings, brushes and brooms, toys and games, sports equipment, stationery, etc.

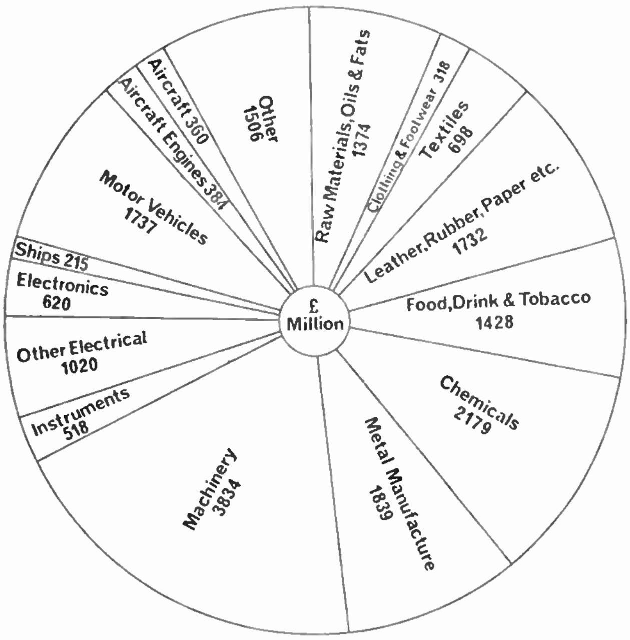


Fig. 4. UK exports 1975. (Total = £19762M)
Source: Overseas Trade Statistics December 1975.

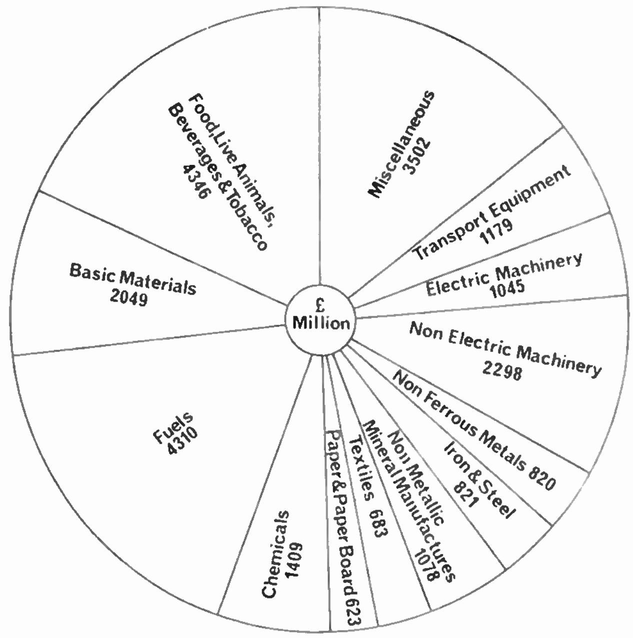


Fig. 5. UK imports 1975. (Total = £24163)
Source: Overseas Trade Statistics December 1975.

added to the value of every ton of coal is worth more than £100 added to each ton of diamonds'.

Imports and Exports

Figure 4 shows in the form of a pie chart the make-up of our exports where it can be seen that machinery, chemicals, metal manufacture, leather, rubber, paper and motor vehicles play a leading role. By comparison aircraft, aero engines and ships are relatively minor contributors. Once more it can be seen that it is the impact of technology on the big sectors that is really relevant to the balance of payments.

The corresponding composition of imports is shown in Fig. 5 where it can be seen that just over a half of the total is composed of necessary raw materials to provide food, basic materials for our industries and for energy. The very prominent item of the import of fuels will be largely eliminated by the development of North Sea oil resources—at least for a period—and the enhancement of the indigenous Coal Industry. The remainder of the raw material imports will be difficult to reduce to any significant extent but any technology that can lead to more efficient utilization of these imported materials, e.g. by re-cycling, can be of benefit.

Of greater interest are those sectors shown on the right of the chart which are composed largely of manufactured items, i.e. products and materials where most of the value has been added by overseas manufacturing industries. The fact that nearly half of our import bill is composed of these manufactured items is an index of the degree of the 'competitiveness' of the home-based industries. If, for example, non-electric machinery or transport equipment is imported, this implies that competitive products (measured in suitability for purpose, price, delivery, etc.) were not available from indigenous manufacturers.

It is worth pursuing this matter further by examining a few of the areas where imports have taken over a substantial part of the home market. I take as my examples the imports of manufacturing machinery. Figure 6(a) shows the level of fixed capital formation (the monies invested by industry in plant and machinery, etc.) for the period 1962 to 1974 and the shaded areas show that portion which was imported. Figure. 6(b) converts these figures into the percentage of imports included in the new plant and machinery installed and it can be seen that from a level of only 18% in 1962 the imported proportion has grown to over 50% in recent years. The drop in the proportion in 1974 should not be regarded as signalling a change in trend; this was largely due to the low overall investment figures for the year (the numbers in Fig. 6(a) have no correction for inflation).

Figure 7 shows the value of the UK market for electrical appliances and the shaded areas once more represent imports.

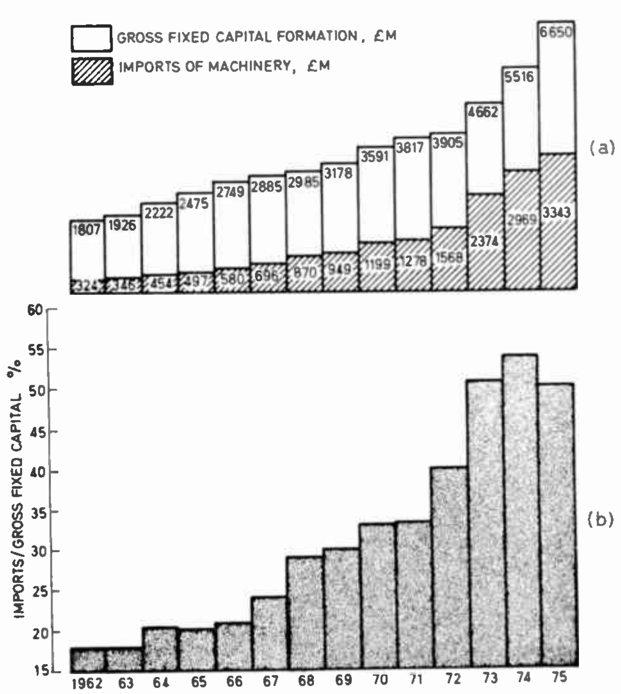


Fig. 6. Imports of machinery and fixed investment.
Source: Economic Trends 1976, Overseas Trade Statistics December 1975.

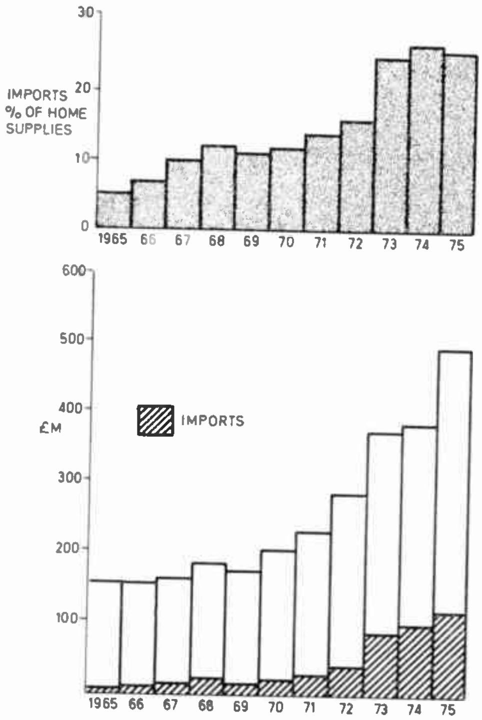


Fig. 7. UK market for electrical appliances deliveries (at current values each year).
Source: AMEDA.

The upper chart gives the percentage figures for imports and shows that the imported proportion has increased from about 5% in the mid-1960s to around 28% in recent years.

What factors are at work here? Classically one looks to such matters as price, availability, customer appeal (looks, timeliness, relevance), quality, reliability, etc. All of these are present in some measure, but price should not have been a factor in view of the sliding exchange rate of the pound. Alas, home costs have been rising at a rate higher than the exchange rate has been slumping and the apparent international price advantage has been eroded. It is relevant here to remind ourselves about the way that the exchange rate has been changing in recent years. Figure 8 gives the mean exchange rate of the pound sterling against the German Mark and Swiss Franc over the past fifteen years. Up until the mid-1960s the value changed very little, though the pound had come under pressure on occasions. Thereafter a steady slide has occurred with the present values standing at appreciably less than half of what it was ten years ago (around 38%). The ratio has changed similarly in relation to other major currencies yet we are still experiencing major incursion into the domestic market in the very product ranges that a high investment in science and technology should have led to international leadership.

The dangers are evident. If the economy is allowed (or encouraged) to expand rapidly between one-third to a half of the more buoyant home market will be filled with imported machinery, motor vehicles, consumer goods etc. If it does then the balance of payment problems will be aggravated further with renewed pressure on the pound etc, etc. The only long-term cure is greater competitiveness of the home-produced products, both in the domestic market and overseas.

Improving Competitiveness

If there were simple routes to improvement in competitiveness these would have been applied long since. A change in national attitudes, in the system of rewards, the

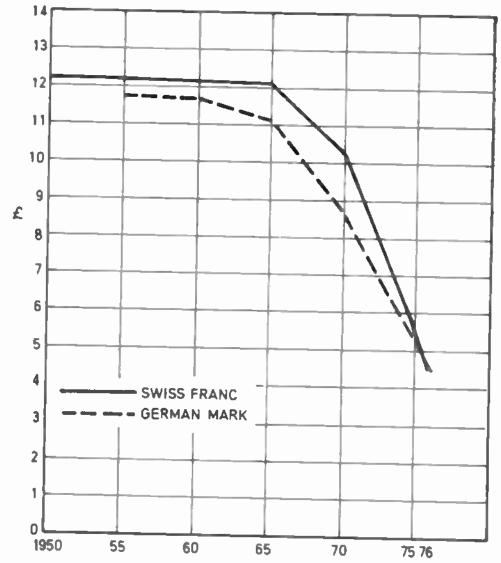


Fig. 8. Exchange rates.
Source 1950–1975: Annual Abstracts 1960 & 1975.

standing and prestige of the manufacturing industries, overmanning, education, training, etc. can all be invoked as elements in the curative process. I will not attempt to pursue these but just concentrate on my central theme that whatever the complex social, political, economic and educational changes that may be necessary they must end up by the more efficient application of technology to our industries. In my view there are two areas where technology can and must play a dominant role. One is in increasing productivity, the other is in improving quality. There is evidence that in both these essential qualities the UK industry—particularly the very large proportion that is based on engineering—has slipped behind its competitors. Dr. F. E. Jones[3] has made comparisons between the industries in the UK and in Japan and he concludes that 'the output per employee in the UK manufacturing industry was only 44% of that in Japan' but goes on to point out that 'the output per £ of plant and machinery was 6% greater (in the UK)' and that 'the added value per £ of total assets employed was 36% greater (in the UK) than in Japan'. From this Dr. Jones concludes that there is a great deficiency of capital equipment. In a more detailed analysis in a later part of his paper Dr. Jones focuses on the productivity of the UK motor industry (i.e. BLMC) and concludes that output per employee was only 37·2% of its Japanese counterpart.

In the Boston Consulting Group report on the motor cycle industry[4] the difference in productivity is even more stark. In exhibit 21 of that report the BCG shows that the UK industry was producing between 10 and 18 motorcycles per man-year out of a typical annual output of a few tens of thousands whilst the Japanese figures were in the region of 150 to 200 per man-year out of a typical annual output of one to two million per year.

Similar adverse ratios have been produced for steel-making, shipbuilding, electronic consumer goods, domestic appliances, machine tools, etc. It has to be faced that our productivity is now low compared with our competitors overseas and much, but not all, of this can be traced to inadequate supporting capital equipment. This is ironic in a nation that built its strength by creating and then using advanced (for the time) manufacturing machinery.

I have long been convinced that we must:

(a) Greatly increase our fixed capital investment—but this *must* be in the most appropriate and timely technology, deployed in an efficient way. New machines scattered at random over an assortment of manufacturing enterprises without cohesion will be of litttle benefit.

(b) Make very full use of these new capital resources. This will call for radical changes in outlook, the willingness to close down old and outmoded plant, slim down manning levels, alter the balance of skills, shift work, etc. Above all ensure that a high quality of 'technological management' is used in our factories calling for the very ablest people, properly educated and trained for the task and, it must be said, properly rewarded.

(c) Rationalize our products and our manufacturing resources to make more capital-intensive manufacturing viable.

Until we are prepared to do these things, our high national investment in science and technology will continue to be abortive.

Quality

It is much more difficult to get any quantitative assessment of quality in manufactured goods. There is however a great volume of anecdotal evidence which points to the fact that many parts of the UK industry fall short of their international competitors in the question of quality. Quality is largely a value judgment but can be rationalized into the following four ingredients.

(1) *Initial Quality*. Is the product well conceived and well executed in its detail? The last point cannot be emphasized too strongly. Clever conceptual designs can be completely vitiated by casual or even erroneous detail designing. The Feilden Report of the early 1960s pointed to this danger[5] but insufficient progress has been made to raise our standards of detail design. We need to lower our sights from the creation of dazzling new concepts and attend much more to rigorous and elegant design of detail. Almost the whole range of Japanese products is orthodox and conventional in concept but it scores in its attention to detail.

(2) *In-service Performance*. Customer satisfaction is rapidly eroded if the product fails in some vital way early in its working life or fails persistently throughout it. Even the failure of ancillary and non-essential features (such as detaching handles or knobs, corroding embellishments, etc.) can lead to the disenchantment of the customer. To a country dedicated to high science and high technology these may appear to be subjects unworthy of attention, yet the route to success lies here.

(3) *Maintainability*. Even the best conceived and engineered products will fail at some stage and the ability to service and replace components is essential to the user. The product must be designed with this in mind, removal and replacement of wearing or fatiguing parts must be easy and spares must be readily available. This calls for close contact between the marketing and design activities of the company and this is made easier if the marketing team includes a significant number of highly able engineers. We do not seem to produce marketing engineers in the UK.

(4) *Specification and Quality Assurance*. In an ever-increasing range of products, it is necessary to work to specifications, particularly if the product has to interact with others in some form of complex. The production of specifications is an essential companion to producing improving technologies. But such specifications need some means of verification and this calls for credible and well-equipped (in both people and plant) quality assurance organisations.[6] The combined use of specification and Quality Assurance Service can do much to improve quality and call for more attention than they currently receive.

'Learning' (Experience) and 'Forgetting' Curves

Anyone with experience in running large, long duration projects will be familiar with the learning curve phenomena. In the initial stages of the project everything is new, innovative or exploratory. Errors are numerous, blind alleys are explored and abandoned, concepts need to be revised and the basic elements have to be discovered or established. As the project develops the most relevant skills emerge and develop, irrelevancies are shed, optima are discovered and management disciplines are established. Progressively the whole process becomes more streamlined, more certain and more efficient. The more complex the project is, the more profound is the effect of this learning curve process. It is not surprising therefore that those who have painstakingly and successfully climbed a long learning curve are reluctant to start again at the bottom of some curve. I am afraid that this phenomenon is beyond the comprehension of many of our inventors, particularly those in academia, who seem to expect organizations that have struggled hard for many years to climb a particular curve to shed all this benefit in order to pursue some new, unproven idea. Frequent jumps into new technologies will mean that the organization will be constantly at the bottom of a succession of learning curves and there is reason to believe that we have done too much of this in the years since the war.

The Boston Consulting Group have pointed to an analogous phenomena (and they use it as a part of their management consultant techniques) which they call experience curves. They have shown that as the aggregated volume of manufacture of a product increases (e.g. production of a particular type of television set) there is a steady and *significant* improvement in unit costs due to accumulated experience. Unit costs continue to decrease for a long time (which may be as much as a decade) and costs follow a power law relationship to the accumulated totals produced. This is a benefit of high (and sustained) output which is over and above the other volume dependent effects such as distribution of overheads, bulk purchasing, specialization etc.

Ultimate success in the more modest technologies can depend more on the speed and persistence of the learning curve than on the generation of new concepts. As the level of education of many previously under-developed nations improves their ability to climb learning curves is increased. Most technologies are now internationally available, through published works, conferences, courses, etc, and all over the world new competitors are emerging furiously climbing their learning curves. The success of Japan, Singapore, Hong Kong, Taiwan, Korea, etc., in mastering quite advanced technologies in a relatively short time is an example of this. The newcomer has the advantage of a 'green field'. He has no preconceived notions or fixed attitudes, he can choose his technology from the best source, frequently producing an advantageous amalgam from several sources, he can focus his attention on the most important part of the market and he can build up on a rational basis.

This brings me to the 'forgetting curve'—a phenomenon that has received very little study. When a company or indeed a country has been a pioneer it is liable to have established practices and attitudes that later become a liability. The resources, the training, the management approach of the past may no longer be relevant. The problem becomes not so much the speed at which the new learning curve can be climbed but the speed at which old habits and resources can be shed. Many of our industries have been caught between the jaws of this learning–forgetting curve interplay. We have too many cases where the old Irish saying could be used, 'If you want to go there I wouldn't start from here'.

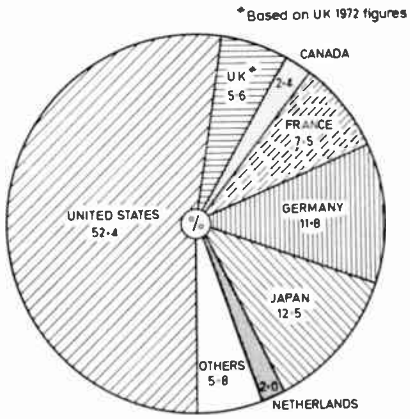


Fig. 9. OECD countries national expenditure on scientific R&D —1973.
Total all OECD Countries £57·9 billions = 100%.
Source: OECD DSTI/76/5050, EEC CREST/65/75.

In the world of industrial technology I see not so much a competition between new and original concepts and techniques as a competition between the rates of learning and forgetting curves. Forgetting curves seem to be longer, more painful and more disruptive than learning curves and the advantage lies with the newcomer. The UK must wake up to the fact that it is now just one relatively small nation, competing in an international arena occupied by an ever-growing number of new learning curves. We will have to be willing to learn and, alas, forget at the same speed as our competitors if we are to survive.

Sources of Technology

There is very little evidence that our high national preoccupation with research has produced corresponding industrial benefits. Indeed it could have been counterproductive in the sense that it breeds an unwillingness to accept the fruits of other nations' R&D and an antipathy to other people's technology. Figure 9 shows in the form of a pie chart the distribution of R&D investment in the OECD countries in 1973. It shows that large though our national investment might be it is only about 5% of the total of the countries involved. Put another way 95% of the R&D in the non-communist world is done *outside* the UK and to blinker ourselves to the extent of denying ourselves access to this 95% is damaging. We have been the most reluctant importer of foreign technologies in the form of licence, know-how agreements, etc., yet when we import foreign-made products we are, *inter alia*, importing foreign technology.

One measure of our reluctance to import foreign technology, compared to our international competitors is the ratio of royalty transactions for imported and exported technology. Table 2 shows these ratios for Japan, Germany and France for the period 1964–1974 and of special significance are the comparisons for 1964. A substantial portion of the current Japanese, French and German competitive products derives from the technology they absorbed from foreign sources at (and before) that time. The fact that the Japanese and French ratios have dropped in more recent times is because many of the early patents have begun to expire—but the products continue.

Table 2. Technology Royalty Transactions 1964–1974

Ratio: Receipts to Purchases	Japan	Germany	France	UK
1964	1:24	1:2·4	1:2·5	1:1
1969	1:8	1:2·5	1:1·7	1:1·1
1973	1:4·5	1:2·5	1:1·5	1:1·1

Nationally we have no prospect of being leaders in every possible field of technology—no matter how distinguished our R&D may have been—and we must learn to import foreign techniques and methods just as we have all too readily learned to import their products.

Putting Techology to Work

I return to the title of this paper and give my conclusions on the measures that are essential if we are to derive increased benefit from our high national investment in R&D in the past. These are presented in roughly descending order of priority.

(1) Increase our capital investment, but in doing so paying particular attention to the technologies we are introducing.

(2) Equip industry with high quality 'techno-managers' but these should *not* be drawn from the ranks of people who are regarded as 'not good enough to do research or advanced design'. They should be the ablest people in the entire organization.

(3) Attend to quality control and improve and expand our quality assurance services.

(4) Increase our emphasis on detail design and elevate the status of this work to the same level as conceptual design.

(5) Be more prepared to learn from other nations.

(6) Appreciate the importance of learning curves (or experience curves) and stop chopping and changing.

(7) Increase our awareness of 'technological marketing' and ensure that engineers of high calibre are directly involved in the marketing operations—and give these engineers strong influence on the Research—Development —Design activities of the company.

(8) Temper our national dedication to advanced (and usually 'big') research with the realization that success comes from humbler but more vital activities.

This pecking order will surprise many and will certainly offend the nation's large army of researchers. But I have become more and more convinced during the past decade that we must become far more concerned with applying technology rather than generating it. We have been so busy assaulting the Technological and Scientific Everests that we have left the more fertile valleys for others to till and to graze.

References

1. 'Member Countries Performance in Technological Innovation', Vol. III of 'Gaps in Technology Between Member Countries: Analytical Report' (OECD, Paris, 1969).
2. Maddock, I., 'End of the glamorous adventure', *New Scientist*, **65**, p. 375, 13th February 1975. Also in *National Electronics Review*, **11**, No. 4, p. 80, July/August 1975.
3. Jones, F. E., 'The economic ingredients of industrial success' (James Clayton Lecture), *Proceedings of the Institution of Mechanical Engineers*, **190**, No. 16, p. 115, February 1976.
4. Boston Consulting Group 'Strategy Alternatives for the British Motor Cycle Industry' HMSO, 1975.
5. The Feilden Report on Engineering Design, HMSO, 1963.
6. Allaway, P. A., Presidential Address, *The Radio and Electronic Engineer*, **47**, No. 1, January 1977.

Manuscript received by the Institution on 22nd September 1976 (Address No. 51).

IERE News and Commentary

Computer Systems and Technology Conference

The Institution of Electronic and Radio Engineers in association with the IEE and B.C.S. are mounting a three day computer conference in the spring of 1977. The University of Sussex has been chosen as the venue for the conference which will take place from Tuesday 29th March to Thursday 31st March 1977. It will be residential.

Professor Maurice Wilkes, F.R.S. will present the keynote address and papers will be grouped under the headings: Microprocessors (systems, design, applications, distributed and dedicated systems), Reliable Systems, and Interactive Systems. (See November Journal, page 569.)

Further details may be obtained from the Conference Department, IERE, 8–9 Bedford Square, London WC1B 3RG. Tel. 01–637 2771.

CEI Chairman Writes to P.M. about Engineers in Industry

Stating its 'grave concern' at the fall-off in the intake of 'quality' young people to industry, CEI has offered its help in any action Mr. Callaghan might initiate to solve the problem. In a letter to the Prime Minister, CEI's Chairman, Mr. Tony Dummett, has stressed that the Council, in conjunction with its constituent members and others in the ERB, is responsible for the qualification and registration of professional engineers. 'As such', says Mr. Dummett's letter, 'we are deeply involved not only in the whole question of education for, and recruitment to, the engineering profession, but with all levels of technical competence'.

The Chairman continued by saying that for some time past CEI has been 'gravely concerned at recent developments whereby the intake of young people into industry has fallen off to some extent in numbers but much more seriously in quality. Indeed it seems clear that we shall not get the standard of management and engineering competence that we need in industry, whether public or private, until an industrial career becomes one of the first choices of our brightest school leavers. All professions are involved, of course, but engineers will form the largest proportion of them'.

After offering the Council's help in any action that the Prime Minister might be initiating, Mr. Dummett concluded by saying that much of the action initiated from various quarters had been uncoordinated. 'It seems to us that there is an important job of co-ordination to be done here if effort is not to be wasted.'

Vacation School on the Practical Use of Control Theory

This Easter Vacation School, to be held at Reading University, on 22nd and 23rd March 1977, on the Practical Use of Control Theory, represents a new concept in postgraduate training. It is designed to bring participants together with practising experts who are also experienced teachers who will describe in a short series of lectures and discussions how to use Control Theory in the solution of a variety of practical problems met in control engineering. The lectures will be entitled:

Review of Control Theory and its Applications
The Root Locus Method and its Applications
Plant Modelling in Interactive Control System Design
Development in Interactive Computer-aided Design of Closed-loop Control Systems
The Design of Tracking Servomechanisms
Design of Missile Autopilots
Applications and Design of Practical Servomechanisms

The School will consist of three sessions—afternoon, evening and the following morning. The fee will be £15 and overnight accommodation in a Hall of Residence will be available at an extra charge of about £8.30.

This first School will be limited to 30 participants but a wider audience, invited from the Thames Valley Section, will join in the evening session.

Further details may be obtained from Mr. P. Atkinson, Department of Engineering and Cybernetics, University of Reading, Whiteknights, Reading RG6 2AL (telephone: 0734 85123) or from the Institution.

EITB Appointment

The Engineering Industry Training Board has appointed Mr. Chris Carroll, C.Eng., F.I.Mech.E., to be the manager of its Electrical & Electronic Sector, responsible for implementing the training policies of the Board within 1,200 companies employing some 550,000 people. Mr. Carroll is a professional engineer with a research and development background in the aircraft and electronic industries, BAC and English Electric respectively. He joined the EITB in 1966 and since 1973 has been Regional Training Officer for the Midlands.

T. Eng. Registration

The Engineers Registration Board has approved in principle a simplified procedure to replace the annual issue of renewal slips for insertion in the registration card.

It had been hoped to implement this procedure during 1976 but this was not possible. The change will, however, take place during 1977 and so renewal slips for 1977 will not be issued.

Associate Members should therefore cut out the slip below and insert it in their registration card.

Because of proposed changes in the procedure for registration renewal, the validity of the renewal slip for 1976 is extended to 31st December 1977.

T. Eng.(CEI) IERE

The Births of Television

MAURICE EXWOOD, C.Eng., F.I.E.R.E.

By some remarkable coincidence we celebrated in 1976 three significant milestones in British Television:

In January the 50th anniversary of the first public demonstration of television, by Scotsman John Logie Baird.

In September the 21st birthday of the start of Independent Television from London.

In November the start of 'high definition' television by the BBC from Alexandra Palace, 40 years earlier.

Both BBC and ITV gave us ample opportunity to indulge in the favourite pastime of the old: nostalgia. Many of us loved the chance to remember and be reminded of the early days, when the box in the corner of the room was not yet taken for granted. Many will have recalled the anxious anticipations of our exposure to something new and naughty: the first commercial, and wondered whether that iceblock was so difficult to recognize at the time. Perhaps it was, for one thing incidentally demonstrated by those 'yesterdays' of 1955 was the enormous advance in picture quality and in particular telerecording, for it was not until 1958 that the first video tape recorders arrived in this country.

Some of us remembered the great changes in receiver technology and the frantic efforts by ITV, rental companies and dealers to get sets converted to Band III by the fitting of a convertor. For most sets in use then tuned only to one of the 5 channels in Band I, indeed some may still have been in use that were fixed tuned to one station! The convertors proved a makeshift solution, but served their purpose in helping to give ITV companies badly needed viewers. But receivers featured little in the commemorative programmes, for although they represent a far larger investment than studios, links and transmitters together and at least equal engineering ingenuity, they are, as Dr. Allaway our president would probably agree, taken as much for granted as the toaster and the 'fridge'. But I for one was fascinated by what Marsland Gander, 'the world's first television critic and correspondent' recalled[1] about the sets pre-1939. The monster dual standard set, Baird 240 line, E.M.I. 405 line plus radio bands and gramophone facilities must have been quite something. His belief that 16 000 sets only were in use in September 1939, in spite of price cuts from £100 to as low as £23, confirms my view, that on the basis of set count, public acceptance of BBC high-definition television in 1939 was not much greater than that of Baird's 30 line, which could be seen by 5000 sets in the early '30s.

The BBC's celebrations of their jubilee culminated in the screening of the monumental production: 'The Birth of Television', a well balanced mixture of documentary and entertainment, broadcast on November 1st. Having tried my hand researching the early days of television for my IERE monograph on Baird,[2] I can only admire the BBC researchers and producers for digging out so many facts, bringing to the screen so many early personalities and making it all come to life.

In terms of publicity by the mass media, the golden jubilee of Baird's historic demonstration came off much the worst. With the exception of *The Guardian*,[3] neither newspapers nor broadcasters treated it as a significant historic event. This is a pity because in my view we can be very proud of Baird, the British inventor, who was the first to make television work, first to broadcast television programmes and first to bring about a regular public television service by a BBC transmitter. He had many other 'firsts' to his name[2], but these will do for starters.

Several researchers marked the jubilee by publishing their findings in the technical press[4-7] and one of these has aroused a deal of controversy in the correspondence columns of *Electronics and Power*.[9] We can assume that the last word has by no means been written about John Logie Baird.

The BBC documentary tried hard to be objective and did much to recognize the pioneering work of Baird and his associates. It established that broadcasts of Baird television started in September 1929 over 2LO, the BBC London station. That is nearly 3 years earlier than the only mention of Baird, and the first of television, in the list of BBC dates in BBC Handbooks so far published. But no viewer could have guessed from this programme that the world's first broadcast television transmissions started in fact another 3 years earlier via Baird's own transmitter in Central London, 2TV (250 watts compared with 3kW for 2LO) in September 1926 or thereabout.

For me the tributes to Baird and his work by Ben Clapp, Tony Bridgewater, Margaret Baird and, in his closing words, Leslie Mitchell, were at least in part offset by the somewhat cynical treatment by BBC people in dealing with Baird. They were recording history when making it clear that the BBC, in the late '20s, disliked Baird and his methods. But it seemed to me to lack objectivity to stress and hold up to ridicule the primitiveness and limitations of the early mechanical system, particularly since this appeared to be done in an effort to justify the ultimate abandonment of the Baird system on which point the documentary showed the BBC highly sensitive and defensive. Quite unnecessarily so: by the time of the opening of Alexandra Palace, Baird & Co. had had 10 years to develop a system with standards acceptable to the public and comparable with those of other visual media at the time, such as films. But they had chosen to spend most of the money publicly subscribed on extravagant promotion of the 30 line version, not only in Britain but in the USA, Australia, Germany, France and other European countries, and little money on attracting and encouraging scientists and engineers to develop a better system. Therefore the BBC had the clear duty to recommend that the insufficiently-improved system of the Baird Company be abandoned in favour of EMI's, after both had a reasonable trial. The Baird results in that trial could have been adversely affected by the disastrous fire 4 weeks after the start, when Crystal Palace and with it all Baird's equipment and records were destroyed. A possible factor not mentioned in the programme.

The BBC's and particularly Eckersley's efforts to deny Baird the part-time use of the under-utilized transmitters in 1927–29 are less easy to justify (2LO broadcast for less than 70 hours per week in 1927) and the Post Office were quite right to bring pressure on the BBC to get them to agree to the use of a transmitter for television outside broadcasting hours. It was Eckersley who set the tone for the BBC attitude towards Baird's efforts by making them look ridiculous. To continue to do so 50 years later seems to me undignified for surely it was a great achievement that there was any television at all and we should be very proud that it started in Britain years before any other country. Certainly Baird's mechanical system was not thought ridiculous by all: not until January

1934 did EMI abandon the idea of a mechanical camera and as Jesty points out,[10] a 405-line mechanical film scanner was developed by GEC in the late 'thirties.

To be fair we should acknowledge that, after Eckersley's resignation in 1929, Baird had a lot of help and encouragement from the BBC, in particular from Sir Noel Ashbridge who succeeded Eckersley as Chief Engineer.

In spite of recently published researches on Baird's work, a number of mysteries remain for me. For instance, how much help did Baird have in the early days? Ben Clapp recalling in 'The Birth of Television' that he joined Baird in 1926, stated that Baird did it entirely by himself. From various contemporary publications I had come to the same conclusion. But Waddell,[8] who did a great deal more research than I, shows that two employees of Will Day (a cinema operator in London and Baird's first non-family financial backer) made the equipment used for the demonstration in Selfridges in April 1925, to Baird's specification. Also in the May 1976 issue of *Electronics and Power* V. R. Mills comes forward (as far as I know for the first time) to state that Baird could not get his early equipment (in Hastings) to work and called on him for help with the result that . . . 'together in 1923 we did succeed in transmitting an image which was moving and recognisable'. These apparent contradictions warrant further investigation.

I also remain mystified as to the light-sensitive devices he used and where he got these or whether he made them himself. He apparently had abandoned selenium by January 1925 and was using a 'colloidal (fluid) cell of my own invention'.[11] He does not seem to have fulfilled the expressed hope to give particulars at a later date.

So uncertainty remains as to the details of what Baird achieved, but all the new material that I have studied has not changed my view that Baird was a great inventor and should be honoured as such.

References

1. L. Marsland Gander, 'When "the box" was opened 40 years ago', *Daily Telegraph*, 1st November 1976.
2. Maurice Exwood, 'John Logie Baird: 50 years of Television' (IERE, January 1976).
3. Peter Fiddick, 'Fifty years of vision', *The Guardian*, 26th January 1976.
4. R. W. Burns, 'The first demonstration of television', *Electronics and Power*, **21**, No. 7, pp. 953–6, 9th October 1975.
5. P. Waddell, W. V. Smith and J. Sanderson, 'John Logie Baird and the Falkirk transmitter', *Wireless World*, **82**, pp. 43–6, January 1976.
6. G. Shiers, 'Television 50 years ago', *Television*, **16**, No. 1, pp. 6–13, January/February 1976.
7. 'A landmark in telecommunications', *Endeavour*, **35**, p. 50, May 1976.
8. P. Waddell, 'Seeing by wireless', *New Scientist*, **72**, pp. 342–4, 11th November 1976.
9. Various correspondents, 'The invention of television', *Electronics and Power*, **21**, p. 1193, 11th December 1975; **22**, p. 41, January; pp. 113–4, February; p. 246, April; p. 310, May 1976.
10. L. C. Jesty, 'The invention of television', *Electronics and Power*, **22**, p. 246, April 1976.
11. J. L. Baird, 'Television—a description of the Baird system by its inventor', *Wireless World and Radio Review*, **15**, pp. 533–5, January 1925.

Addenda and Corrigenda to Reference 2

A number of people have kindly given me information establishing facts which differ from my assumptions (based on early biographies):

(a) Dr. P. Waddell has established that Baird and Hutchinson were apprentices at the Argyle Motorworks, Glasgow, in 1911, years before they agreed to merge their soap business (see p. 3)

(b) He has also drawn my attention to a report in the *Daily Express* of 8th January 1926 which states that Baird and Hutchinson had already extended an invitation to members of the Royal Institution, but apparently no date for the demonstration had been fixed by then.

(c) I am informed by Dr. H. R. L. Lamont that another plaque was placed on the stable block of Kingsbury Manor, North Wembley, in 1953 commemorating the reception by Baird of television signals from Berlin in June 1929 (p. 29)

(d) Reference 31 should read: *Proceedings of the IEE*, **99**.

Standard Frequency Transmissions—October 1976

(Communication from the National Physical Laboratory)

October 1976	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)	Relative phase readings in microseconds NPL—Station (Readings at 1500 UT)	
	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	−0·2	OFF AIR	610·3
2	−0·2	FOR	610·2
3	−0·2	WHOLE	610·2
4	−0·2	MONTH	610·2
5	−0·1		610·2
6	0·0		610·2
7	0·0		610·0
8	0·0		609·6
9	0·0		609·6
10	0·0		609·4
11	0·0		609·5
12	0·0		609·5
13	0·0		609·7
14	0·0		609·9
15	0·0		609·9
16	0·0		609·8
17	0·0		609·8
18	0·0		609·8
19	0·0		OFF AIR
20	0·0		,,
21	0·0		,,
22	0·0		,,
23	0·0		,,
24	0·0		,,
25	0·0		609·8
26	0·0		609·8
27	0·0		OFF AIR
28	−0·1		,,
29	0·0		609·8
30	0·0		610·0
31	−0·1		610·0

All measurements in terms of H-P Caesium Standard No. 344. agrees with the NPL Caesium Standard to 1 part in 10^{11}.

* Relative to UTC Scale; (UTC_{NPL}–Station) = +500 at 1500 UT 31 December 1968.

† Relative to AT Scale; (AT_{NPL}–Station) = +468·6 at 1500 UT 31 December 1968.

Applicants for Election and Transfer

The Membership Committee at its meeting on 18th November 1976 recommended to the Council the election and transfer of the following candidates. In accordance with Bye-law 23, the Council has directed that the names of the following candidates shall be published under the grade of membership to which election or transfer is proposed by the Council. Any communication from Corporate Members concerning the proposed elections must be addressed by letter to the Secretary within twenty-eight days after publication of these details.

Meeting: 18th November 1976 (Membership Approval List No. 227)

GREAT BRITAIN AND IRELAND

CORPORATE MEMBERS

Transfer from Graduate to Member

BARNETT, Francis Paul. *Larkhill, Wiltshire.*
CLARK, Kenneth James. *Billericay, Essex.*
CROSS, Alan Frank. *Windsor, Berkshire.*
EMONYON, Richard Andrew. *Bromley, Kent.*
GODDARD, Thomas Edmund. *Newhaven, Sussex.*
McLEOD, Ian James. *Dunnington, York.*
SHAHI, Baldev Krishan. *Maidenhead, Berkshire.*

Direct Election to Member

BALDWIN, Roger Edwin. *Abingdon, Oxfordshire.*
MALHOTRA, Harmohindar Kumar. *Kidsgrove, Staffordshire.*
VERMA, Sujay Kumar. *Totton, Hampshire.*

NON-CORPORATE MEMBERS

Transfer from Student to Graduate

ILLINGWORTH, Graham. *Hitchin, Hertfordshire.*

Direct Election to Graduate

AUSTIN, Trevor Charles. *Sarisbury Green, Hampshire.*
CARMODY, James Christopher. *Dublin.*
EDOHASIM, Onyema Julian. *London.*
LI, Chi Kwong. *London.*
MUNSON, Kelvin Arthur. *Watford, Hertfordshire.*
ROBERTSON, Andrew. *Esher, Surrey.*
SANGHA, Raj Mohan. *Coventry.*

Transfer from Graduate to Associate Member

JONES, Alan Charles. *Hillingdon, Middlesex.*

Transfer from Student to Associate Member

IBIRONKE, Albert Ige. *London.*
KAY, Ernest Rudolf. *St. Albans, Hertfordshire.*

Direct Election to Associate Member

JEFFORD-HORN, Christopher Roderick. *Wimborne, Dorset.*
KONIBAH, Mamudu Chemogo. *Bristol, Avon.*
PEARCE, Michael Andrew. *London.*
READ, Richard John. *Sawbridgeworth, Hertfordshire.*
SEARLE, Barry Charles. *Bromley, Kent.*
STAVROU, Andreas. *London.*
THORPE, David Victor. *Tunbridge Wells, Kent.*

STUDENTS REGISTERED

AINSWORTH, David Charles. *Birmingham.*
ASHWELL, Richard Lewis. *Bishops Stortford, Hertfordshire.*
COLE, Christopher James. *Peterborough, Cambridgeshire.*
COOK, James Anthony. *Chepstow, Gwent.*
DACOSTA, Howard Scott. *Woodford Green, Essex.*
DANIEL, Joseph. *Plymouth, Devon.*
EICHLER, Peter. *Leicester.*
EVANGELI, Nicholas Thomas. *Wylam, Northumberland.*
GUY, Richard. *Leicester.*
HILL, David Stanley. *North Shields, Tyne and Wear.*
HOLLAND, Stephen William. *Newcastle upon Tyne*
KAVANAGH, Donald. *Newcastle upon Tyne.*
KAY, Alan John. *Leicester.*
KONG KHEE FAH, You Ying. *Uxbridge, Middlesex.*
OWENS, Alan. *Longbenton, Tyne and Wear.*
PARTAP, Robert Kamal. *Plymouth, Devon.*
PASIAK, Stanislaw Anthony. *Leicester.*
PENNY, Stephen David. *Shepperton, Middlesex.*
PERERA, Cyril Raomal. *Swansea, West Glamorgan.*
PHILPOT, Michael John. *Reading, Berkshire.*
POTTER, Robert. *Leicester.*
QUASHIE, Curtis Ambrose. *Wellingborough, Northamptonshire.*
RICHARDSON, Julian Howard. *Portchester, Hampshire.*
SPINOSA-CATTELA, Richard David. *Rowlands Gill, Tyne and Wear.*
UDOH, Ine E. E. *Southampton, Hampshire.*

OVERSEAS

CORPORATE MEMBERS

Transfer from Graduate to Member

CROOK, Colin. *Phoenix, Arizona.*
SUNG, Victor Heung Wing. *Hong Kong.*

NON-CORPORATE MEMBERS

Transfer from Student to Graduate

JIBOWU, O. Mofo. *Ibadan, Nigeria.*
LEE, Ting Fai. *Hong Kong.*
POON, Albert Ka-Fat. *Hong Kong.*

Direct Election to Graduate

LEONG, Yew Kay. *Singapore.*
NG, Shiu Hung. *Hong Kong.*

Transfer from Student to Associate Member

CHEN, Tam Kim. *Negri Sembilan, Malaysia.*
PONG, Kwok-Shu. *Hong Kong.*
WONG, Teck Sung. *Singapore.*

Direct Election to Associate Member

OKONKWO, Lawrence Kanayo. *Onitsha, Nigeria.*
STAPLES, Michael John. *Abu Dhabi, United Arab Emirates.*
WRIGHT, James Alan. *Bahrain.*

Direct Election to Associate

RAJAMOHAN, Tanjore Natarajam. *Kitwe, Zambia.*
OLAOSUN, Gabriel Oladejo. *Lagos, Nigeria.*

STUDENTS REGISTERED

CHAN, Ghim Hong. *Singapore.*
CHAN, Kwong Hung. *Hong Kong.*
CHIU, Kin Fai. *Hong Kong.*
CHUM, Chi Hung. *Hong Kong.*
HO, Tat Meng. *Singapore.*
KODAOLU, Olupemi Bolanle. *Abeokuta, Nigeria.*
LAM, Chun Wing. *Hong Kong.*
LAW, Chun Kwong. *Hong Kong.*
LAW, Moon Hung. *Hong Kong.*
LAW, Wing Huen. *Hong Kong.*
LEE, Boon Kuan. *Singapore.*
LEE, Chung Ming. *Hong Kong.*
LEUNG, Chun Sing. *Hong Kong.*
LEUNG, Yiu-Kai. *Hong Kong.*
LI, Pak Kin. *Hong Kong.*
MAK, Chi Keung. *Hong Kong.*
NG, Kwong Tsan Kelly. *Hong Kong.*
PANG, Jing. *Penang, West Malaysia.*
RANASINGHE, Sumathi. *Moratuwa, Sri Lanka.*
SZE-TO, Cheung. *Hong Kong.*
WONG, Ka Man. *Hong Kong.*
YEUNG, Hok Leung Francis. *Hong Kong.*
YONG, Wai Poh. *Singapore.*
YUEN, Pak Fat. *Hong Kong.*

Letter to the Editor

From: P. E. Carr, C.Eng., M.I.E.R.E.

'Fighting for the Professional Engineer'

Your correspondent C. Ridgers, in the October issue of *The Radio and Electronic Engineer* concludes, with reference to Professional Engineers and Trade Unions, that 'Nobody appears to have asked the Professional Engineer (from any institution) what his feelings are'. A poll was conducted by the United Kingdom Association of Professional Engineers, during December 1973–January 1974 of everyone on the latest available mailing lists of the Institutions of Mechanical, Electrical, Civil and Marine Engineers and the Royal Aeronautical Society. More than thirty thousand replies were received to questions relating to

(a) The Code of Ethics

(b) Political Affiliation

(c) Exclusivity to Professional Engineers

The result (summarized) was:

'In circumstances where professional engineers require the support of an organization having Trade Union Powers, 60% or more considered that ethical behaviour, non-political status and professional exclusiveness were essential. Over 94% rated the same requirements as preferable, important, or essential. 95% expressed the view that professional engineers should not be expected to join a trade union against their wishes'.

The analysis was carried out by M. P. Newman, C.Eng., M.I.Mech.E.

P. E. Carr

31 Abbots Road,
Abbots Langley,
Watford WD5 0AY.
15th November 1976

Forthcoming Institution Meetings

London Meetings

Wednesday, 19th January

JOINT IERE/IEE COMPUTER GROUP

Colloquium on THE ENGINEER'S APPROACH TO MICROPROCESSOR SYSTEMS

Royal Institution, Albemarle Street, London W1, 2.30 p.m.

Advance registration necessary. For further details and registration forms, apply to Meetings Officer, IERE.

Wednesday, 26th January

MEASUREMENTS AND INSTRUMENTS GROUP

Colloquium on R.F. MEASUREMENTS AND CALIBRATION IN BRITAIN

Royal Institution, Albemarle Street, London W1, 10 a.m.

Advance registration necessary. For further details and registration forms, apply to Meetings Officer, IERE.

Thursday, 27th January

COMMUNICATIONS GROUP

Colloquium on A REVIEW OF DATA MODEMS

Royal Institution, Albemarle Street, London W1, 2.30 p.m.

Advance registration necessary. For further details and registration forms, apply to Meetings Officer, IERE.

Tuesday, 8th February

JOINT IEE/IERE/BES MEDICAL AND BIOLOGICAL ELECTRONICS GROUP

Physiology for engineers—4

Botany Lecture Theatre, University College, London, 6 p.m.

Wednesday, 16th February

CEI MEETING

Open Forum on PROFESSIONAL ENGINEERS AND UNIONS

Institution of Mechanical Engineers, 1 Birdcage Walk, London SW1, 6 p.m.

Tuesday, 22nd February

COMMUNICATIONS GROUP

Colloquium on DIGITAL ENCODING OF AUDIO SIGNALS

Royal Institution, Albemarle Street, London W1, 2 p.m.

Thames Valley Section

Thursday, 20th January

Electronic dashboard instrumentation

By P. N. Thomas (*Smiths Industries, Witney*)

Caversham Bridge Hotel, Reading, 7.30 p.m

Wednesday, 23rd February

Signal processing in conjunction with the H-P 3000 computer

By D. Cox (*School of Signals, Blandford Forum*)

School of Electronic Engineering, Arborfield, Berks., 2.30 p.m.

East Anglian Section

Wednesday, 19th January

JOINT MEETING WITH IEE

Recent technological advances in microprocessors

By N. Carruthers (*Digital Equipment Co. Ltd.*)

King Edward VIth Grammar School, Broomfield Road, Chelmsford, 6.30 p.m. (Tea 6 p.m.)

Thursday, 27th January

JOINT MEETING WITH IEE

Digital television—a logical choice

By K. H. Barnett (IBA)

University Engineering Laboratories, Trumpington Street, Cambridge, 6 p.m. (Tea 5.30 p.m.)

South Western Section

Tuesday, 8th February

The history of sound recording

By D. Aldous and J. Pengelly

Lecture Theatre, Plymouth Polytechnic, 6.30 p.m. (Tea 6 p.m.)

Wednesday, 9th February

JOINT MEETING WITH RAeS AND IEE

The physical integration of aircraft electronic and instrument systems

By H. Zeffert (*BAC*)

Queen's Building, School of Engineering, University of Bristol, 6.30 p.m. (Tea 6 p.m.)

Wednesday, 2nd March

Acoustical imaging and holography

By Professor J. W. R. Griffiths (*Loughborough University of Technology*)

Room 2E3.1, Bath University, 7 p.m. (Tea 6.30 p.m.)

Southern Section

Wednesday, 19th January

JOINT MEETING WITH CEI

Radar assisted berthing of large tankers

By S. Thompson (*formerly Marconi International Marine Co.*)

Lecture Room 4A, Mathematics Building, Southampton University, 7 p.m.

Synopsis: With the increasing use of very large tankers for the transportation of crude oil, the accurate measurement of the tankers' speed of approach to a land-based terminal is vital. The equipment found on such vessels is not usually sufficiently accurate and several alternative techniques have been considered. Of these, a Doppler radar system is the most suitable and a description of these techniques will be presented. Aspects of the problem from the point of view of the Master and Pilot on board and of the Berthing Master ashore will also be discussed.

Wednesday, 9th February

Radar applications of surface acoustic wave devices

By A. Thompson (*Plessey Radar*)

Portsmouth Polytechnic, King Henry 1st Street, Room ABO-11, 7 p.m.

Synopsis: The versatility of surface acoustic waves (SAW) has produced a variety of devices which have found application in radar systems. The successful development of these devices has been mainly due to the planar structure of SAW technology which permits interaction with the wave at any point in the acoustic path. The ease with which the wave may be sampled is particularly important in signal processing. The items to be described will include matched filters, code generators and oscillators.

Friday, 11th February

Radio astronomy

By M. C. Kemp (*Mullard Radio Astronomy Observatory, Cambridge University*)

Isle of Wight College of Arts and Technology, 7 p.m.

Thursday, 24th February

Instrumentation in North Sea oil technology

By L. C. Towle (*Measurement Technology*)

South Dorset Technical College, Weymouth, 6.30 p.m.

Synopsis: The special environmental conditions presented by a combination of the North Sea and extracting a hazardous gas or liquid will be outlined. The measurement and control problems met on both exploration drilling and extraction installations outlined and discussed. A more detailed analysis of some more unusual measurements will conclude the lecture.

Tuesday, 1st March

Communications for disabled persons

By A. F. Newell (*Southampton University*)

Bournemouth College of Technology, 7 p.m.

Wednesday, 2nd March

EDUCATION AND TRAINING GROUP

Colloquium on THE FUTURE OF HIGHER DIPLOMAS

Southampton College of Technology, 10.30 a.m.

Kent Section

Thursday, 10th February

Switched mode regulated power supplies

By P. Chapman (*Marconi-Elliott Limited*)

Medway and Maidstone College of Technology, Horsted, Chatham, 7 p.m.

Synopsis: The basic simplicity of the switched mode technique is demonstrated by referring to the design equations of three basic types. Understanding some of the subtleties of these circuits is an engineer's first problem, for the in-depth analysis of large-signal non-linear circuits is by no means easy. Thus, a summary of the features of the basic power circuits and possible control circuit configurations is given. A new set of problems is presented by component selection and system considerations (e.g. EMC). These and other 'real life' problems will be discussed. Finally, a number of less well known but interesting circuits will be presented, and consideration given to the impact of switched mode techniques in areas other than power supplies.

Beds & Herts Section

Tuesday, 18th January

Control and supervision of television transmitters

By V. Arnold (*Granada TV Rentals*)

Mander College, Bedford, 7.45 p.m.

Wednesday, 16th February

Electro-optic effects in liquid crystals

By Ian Shanks (*Royal Signals and Radar Establishment*)

Watford College of Technology, Hempstead Road, Watford, 7.45 p.m.

Synopsis: Electro-optic effects in liquid crystals have received much attention over the last eight years with a view to making high contrast display devices which modulate ambient lighting rather than trying to compete with it by emitting light. Such displays commonly exhibit microwatt power consumptions and can be operated using a.c. r.m.s. voltages as low as 3 volts. The first generation of such displays made use of a phenomenon known as dynamic scattering which involved the flow of ionic charges through a thin liquid crystal layer. The second and present generation use a polarization switching effect called the 'twisted nematic' effect to control the transmission of light through linear polarizers. This requires no ionic current flow and gives a high contrast display usable in digital watches, calculators, digital voltmeters, etc. Recent improvements in liquid crystal materials and display fabrication techniques have allowed the production of displays with commercially viable lifetimes. A third generation of displays is presently at the research stage, and this promises to give brighter and more easily legible displays having no 'Polaroids'. The properties of these different effects will be discussed and a number of novel uses described.

West Midlands Section

Tuesday, 18th January

Some aspects of running a local radio station

By D. Wood (*Chief Engineer, BRMB*)

Department of Electronic and Electrical Engineering, University of Birmingham, Edgbaston, 7 p.m. (Tea 6.30 p.m.)

Wednesday, 16th February

JOINT MEETING WITH ROYAL TELEVISION SOCIETY

Electronic journalism

By Peter Ward (*ITN*)

ATV Centre, Birmingham, 7 p.m.

Thursday, 17th February

JOINT MEETING WITH IEE AND IPOEE

Topics in the early history of telephony

By Professor D. G. Tucker (*University of Birmingham*)

Post Office Technical Training College, Stone, 7 p.m. (Tea 6.30 p.m.)

East Midlands Section

Tuesday, 18th January

8080 microprocessor family application

By Dr. D. J. Quarmby (*Loughborough University*)

Leicester University, 7 p.m. (Tea 6.30 p.m.)

Tuesday, 15th February

JOINT MEETING WITH IEE

Accountancy in the engineering industry

By C. P. Jones and J. R. G. Schofield (*Brush Switchgear*)

Edward Herbert Building, Room J002, Loughborough University of Technology, 7 p.m. (Tea 6.30 p.m.)

South Midlands Section

Tuesday, 8th February

JOINT MEETING WITH IEE

How wrong can you be? Uncertainty and confidence in measurement

By F. L. N. Samuels (*Department of Prices and Consumer Protection*)

Plough Hotel, High Street, Cheltenham, 7.30 p.m.

Wednesday, 16th February

Communications for North Sea oil installations

By L. J. Botton (*P.O. Telecommunications*)

Foley Arms Hotel, Malvern, 7.30 p.m.

North Eastern Section

Tuesday, 11th January

High speed digital transmission

By G. W. Goddard (*Post Office Research Centre*)

Y.M.C.A., Ellison Place, Newcastle, 6 p.m. (Tea 5.30 p.m.)

Tuesday, 8th February

Acoustical imaging and holography

By Professor J. W. R. Griffiths (*Loughborough University of Technology*)

Y.M.C.A., Ellison Place, Newcastle, 6 p.m. (Tea 5.30 p.m.)

North Western Section

Thursday, 20th January

Optical fibre communications

By a speaker from the *Post Office Research Department*

Renold Building, U.M.I.S.T., Sackville Street, Manchester, 6.15 p.m. (Light refreshments available before meeting)

Thursday, 24th February

Pocket computers

By T. A. Self (*CBM Business Machines*)

Renold Building, U.M.I.S.T., Manchester, 6.15 p.m.

South Wales Section

Wednesday, 12th January

Microprocessor systems

By M. Healey (*University College, Cardiff*)

Room 112, Department of Physics, Electronics and Electrical Engineering, U.W.I.S.T., Cardiff, 6.30 p.m. (Tea 5.30 p.m.)

Synopsis: The concept of an integrated circuit microprocessor is now well established. The use of these devices in complete systems is more problematic. This lecture will discuss the salient characteristics of available microprocessors. More attention, however, will be given to the support hardware and software required to build special-purpose equipment. Practical examples of controllers in control and data communications will be discussed, with the accent on how it was done rather than what it does.

Wednesday, 9th February

Electronics training in the Army

By J. B. Helder (*Army School of Electronic Engineering*)

Room 112, Department of Physics, Electronics and Electrical Engineering, U.W.I.S.T., Cardiff, 6.30 p.m. (Tea 5.30 p.m.)

Synopsis: The Army selection system is described and it is shown how potential technicians are selected on aptitude tests, attitudes and interest. Successful recruits are provided with a comprehensive career training and employment and even without 'O' levels or CSE grades, students have finally achieved HNC and T.Eng (CEI) qualifications. Technician career education and training is shown to be a classic 'thick sandwich' of progressive training and skilled employment. Mention is made of the educational technology involved in the design of courses, methods and media and their validation. Continuation training for Officers beyond their first degree is also explained.