

FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

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

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 33 No. 3

MARCH 1967

A New Communication Satellite Earth Station

SATELLITE communication using synchronous satellites, first envisaged by an Englishman, Arthur C. Clarke, in an article in *Wireless World* in 1945, has taken less than 20 years to come to reality, although it was not until the launching of the first *Sputnik* research vehicle in 1957 that the stage was set for subsequent satellites for communications. Progress has been rapid: the passive reflector *Echo I* in 1960, the non-synchronous *Telstar I* in 1962, the synchronous *Syncom II* in 1963, the Russian *Molniya I* and *Early Bird* in 1965. The early doubts expressed regarding the effect of transmission time delay on intelligible telephone conversations have proved to be unfounded and it is apparent that satellites will soon fulfil a full-time role in international communications by providing hundreds of reliable channels.

The use of communication satellites will be further advanced when a new satellite Earth station terminal at Goonhilly, Cornwall, is brought into service by the British Post Office in April 1968. Referred to as Goonhilly II, it will be one of the most advanced satellite communications stations in the world. Ninety feet in diameter, the aerial of this new station will be larger than its predecessor and, with its advanced transmitting and receiving equipment, it will be capable of covering the civil satellite communications bands (3.7-4.2 GHz for satellite to Earth station and 5.925-6.425 GHz for Earth station to satellite).

The station will be able to receive a television signal and up to 500 telephone channels simultaneously—this represents more than five times the handling capacity of the first Goonhilly station. When the first of the next generation of global communications satellites, *Intelsat III*, goes into service next year, its 'multiple access' facility will allow communication with several earth stations simultaneously. The complete system will be wholly in compliance with C.C.I.R. recommendations and the internationally agreed communications performance standards recommended by the Interim Communications Satellite Committee on behalf of the 55 member nations of *Intelsat*.

When the new aerial is completed and takes over the existing transatlantic telephone traffic, the present aerial at Goonhilly, which has proved so successful in operation, will have further equipment added to enable it to take on a new role operating to the new *Intelsat III* satellite over the Indian Ocean. This will enable communication to be established by satellite to Australia, India, Pakistan, Ceylon, Japan and the Far East generally.

The new station, which is being built for the Post Office by The Marconi Company, will have duplicated transmitters, using t.w.t. amplifiers with a saturated power output of 10 kW over a bandwidth of 500 MHz. The low-noise receiving stage will have duplicated liquid-helium-cooled parametric amplifiers of 500 MHz bandwidth, and will be mounted in a cabin directly behind the apex of the paraboloid, thus minimizing the waveguide run to the feed.

In 1961 when the Institution organized the Convention on 'Radio Techniques and Space Research', the papers recording solid achievement were all concerned with research satellites and communications applications were all in the form of proposals. Communication satellite systems are obviously reaching the stage at which a full Conference on engineering realization will be justified.

F. W. S.

INSTITUTION NOTICES

I.E.R.E. Membership Qualification Recognized by Nigerian Government

The Ministry of Establishments of the Government of Nigeria has notified the Institution that Corporate Membership of the I.E.R.E. is recognized as qualifying a candidate for appointment as an engineer in Government Service, with grading in Scale A.

Nigeria has thus fallen into line with the policy adopted by the Civil Service Commission, the Burnham Committee, the B.B.C., and the Armed Services in the United Kingdom; and Government, Public Service Commissions and Broadcasting Establishments in India, Pakistan, New Zealand, South Africa, Ceylon, Ghana, Eire, Israel, Cyprus, and East Africa.

Paris Components Exhibition

The Institution is once again exhibiting the *Journal* and other publications at the Salon International des Composants Electroniques, which will be held at Porte de Versailles, Paris, from 5th to 10th April. The Institution's stand will be in charge of Mr. W. G. J. Nixon (Member), and members visiting the exhibition will be welcomed at the stand.

It is not intended to hold a formal meeting of members in France during the time of the Salon because a larger-scale meeting at a more convenient time in the early Autumn is being planned. Further information about the Council's proposals will be announced in due course.

In addition to the exhibition of passive components, semiconductors, valves and constructional parts, an exhibition of audio equipment will be held in an adjoining hall. Other events of interest to the radio and electronic engineer which are to take place in Paris during this period are listed under 'Conferences and Exhibitions' at the back of this *Journal*.

Stochastic Problems in Underwater Sound Propagation

A N.A.T.O. Advanced Study Institute on 'Stochastic Problems in Underwater Sound Propagation' will be held in Lerici (La Spezia) from 18th to 23rd September 1967. The Institute is sponsored by the Italian Navy.

Scientists and engineers from all countries who work in the field of underwater acoustics or are affiliated with universities, naval laboratories or industrial firms specializing in underwater acoustics are invited to participate in the Institute. Lectures will be given in French or in English by speakers from several countries. The subject of the distortion of acoustic signals during propagation in the water will be discussed under the following headings:

Phenomena of sound propagation in a random medium; Structure of signal noise and reverberation; Detection problems; Oceanography.

Further information may be obtained from Professor Maurizio Federici, c/o U.S.E.A., Via P. Mantegazza 23, San Terenzo (La Spezia), Italy. Offers of papers should be made to Professor Federici without delay. (Final copies of texts are required by 30th June.)

Conference on Radio Receiver Systems

The University College of Swansea is organizing, with the support of the Institution, a Conference on Radio Receiver Systems, to be held in the Department of Electrical Engineering, from 11th to 15th September 1967. The Conference will be concerned with the design philosophy of medium and long distance receiving systems. Sessions will deal with the following aspects:

Broadcast receivers; Transmission medium; Components; Communication receivers; and General aspects.

Further information on the Conference, which will be residential, will be published in forthcoming issues of the Institution's *Journal* and *Proceedings*, or may be obtained from the Conference Secretary, Dr. R. C. V. Macario, School of Engineering, Division of Electrical Engineering, University College of Swansea, Singleton Park, Swansea, Glamorgan.

7th M.O.G.A. Conference

The 7th International Conference on Microwave and Optical Generation and Amplification will be held in the Autumn of 1968 in Hamburg, West Germany. It is being organized by the Nachrichten-technische Gesellschaft im Verband Deutscher Elektrotechniker (The Communications Division of the German Association of Electrical Engineers).

Further information will be available later in 1967, but in the meantime enquiries should be addressed to Dr.-Ing. H. Burghoff, Nachrichten-technische Gesellschaft im VDE, Stresemann Allee 21, 6 Frankfurt on Main S.10, Federal Republic of Germany.

Bound Volumes of *Journal*

The Index to Volume 32 was circulated to Members with the February issue of *The Radio and Electronic Engineer*. Members who wish to have the last six issues of 1966 bound may now send their copies, together with the Index, to the Publications Office, I.E.R.E., 9 Bedford Square, London, W.C.1. Orders for binding should be accompanied by a remittance of 25s., plus cost of postage (5s. 6d. for U.K., and 6s. 6d. for overseas).

The PAL Colour Television System

By

B. J. ROGERS†

Summary: The PAL system and its relationship to the N.T.S.C. system is described. The effects of phase distortion and sideband limitation separately and in combination are analysed. The attributes of various decoding arrangements are compared and the action of the delay line discussed. The possibility of error correction prior to transmission is outlined and finally the consequences following the use of systems with a line sequential component are described.

1. Introduction

The PAL system (Phase Alternation Line) is a further development, by Dr. Walter Bruch of the Telefunken Company, of the N.T.S.C. colour television system that has been used for public transmission in the United States of America since 1953 and, more recently, in Japan. To establish its relationship to the PAL system a brief description of this system is given in the following section. Due to the method of modulating the colouring information in the N.T.S.C. system, certain critical features arise which are of particular importance where the signal cannot be closely controlled, i.e. after leaving the transmitting antenna. The PAL system effectively overcomes these difficulties for very little additional complexity in the receiver and has been chosen for the establishment of colour television services by most countries in Western Europe after more than three years of most searching international tests.

2. The N.T.S.C. System

Starting from the red, green and blue gamma-corrected primary colour signals, E'_R , E'_G and E'_B , provided by the camera or other scanning means, the N.T.S.C. signal consists of the following components:

(a) Scanning synchronization as in monochrome television.

(b) A luminance signal

$$E'_Y = 0.3E'_R + 0.59E'_G + 0.11E'_B$$

(c) Two colouring signals modulated on a sub-carrier

$$E'_I = 0.74(E'_R - E'_Y) - 0.27(E'_B - E'_Y)$$

and

$$E'_Q = 0.48(E'_R - E'_Y) + 0.41(E'_B - E'_Y)$$

(d) A colour synchronizing signal transmitted in the horizontal blanking interval following the horizontal synchronizing pulse.

The luminance signal is proportioned according to the luminance contribution of the three primary colours; it is defined by the properties of the eye.

† Rank Bush Murphy Ltd., Chiswick, London, W.4.

The colouring signals, usually referred to as colour difference signals are such that for the monochrome condition, i.e. $E'_R = E'_G = E'_B$, they have zero magnitude. This situation fulfills sufficiently the principle of 'constant luminance', namely that separate signals carry brightness and colouring information, that having chosen a given normalizing white, no colouring information be carried by the luminance signal, and that no luminance information be carried by the colouring signals. The necessity for gamma correction imposed by the display device transfer characteristic in the receiver limits the degree of constant luminance obtained.

Constant luminance permits monochrome receiver compatibility of the colour signal, since such a receiver can use the E'_Y signal provided that means are found to include the colouring information without undesirable effects either to the colour or monochrome receiver. This is facilitated since the amplitude of the colouring signals is proportional to the amount of colour to be transmitted.

It should be noted here that any signal derived from E'_R , E'_G and E'_B having zero amplitude when these signals are equal may be regarded as a colour difference signal; thus the usual colour difference signals are $E'_R - E'_Y$, $E'_B - E'_Y$, $E'_G - E'_Y$, E'_I and E'_Q .

Since the magnitudes of E'_I and E'_Q define the saturation or intensity of the colour to be transmitted and their ratio defines the hue, i.e. which colours, an alternative and convenient notation may be used. Since only signal voltages are to be used and gamma correction may be inferred henceforth, the E' symbol will be assumed to precede the appropriate variables. If I and Q are considered as vectors as in Fig. 1, the saturation and hue of a colour are the modulus and argument of

$$Q + jI \quad \dots\dots(1)$$

Thus the saturation

$$S = \sqrt{Q^2 + I^2} \quad \dots\dots(2)$$

and the hue

$$\alpha = \tan^{-1}(I/Q) \quad \dots\dots(3)$$

I and Q are transmitted by means of suppressed

carrier modulation upon two subcarriers having a common frequency but with 90° phase relationship. The combination of these two modulated subcarriers, known as the chrominance signal, may be written

$$M = Q \cos \omega_0 t - I \sin \omega_0 t \quad \dots\dots(4)$$

The relationship of this combination to the hue and saturation of the colour to be transmitted may be seen by writing the above equation in an alternative form:

$$M = \sqrt{Q^2 + I^2} \cos \omega_0 t + \tan^{-1} (I/Q) \quad \dots\dots(5)$$

Figure 1 still applies to this signal, but the vectors must be considered to be phasors rotating at $\omega_0 t$. This reveals, by comparison with (2) and (3), that the saturation of the colour is carried by amplitude modulation and the hue by phase modulation.

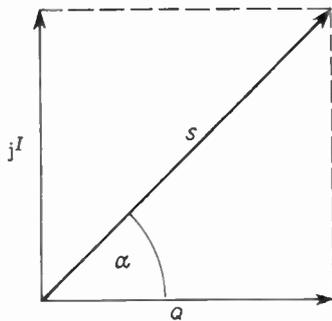


Fig. 1. Relationship of *I* and *Q* components to hue and saturation.

The *I* and *Q* signals are limited to a relatively narrow bandwidth compared with the luminance signal since the eye is much less sensitive to detail where a change in hue or saturation is concerned than is the case for a change in luminance. In fact, the *I* signal corresponds approximately to the gamut of colours in which the eye is most sensitive to detail and the *Q* signal to that in which the eye is approximately the least sensitive to detail. The effective bandwidths of the three signals for the U.K. 625-line standard (C.C.I.R. standard I) are:

$$Y \ 5.5 \text{ MHz}, \ I \ 1.6 \text{ MHz}, \ Q \ 800 \text{ kHz}.$$

The spectrum of a television signal for normal picture content has two significant features: first the energy distribution as a function of frequency is near Gaussian, there being much greater energy content at the lower frequencies; secondly the energy content is in discrete bunches at harmonics of the line scanning frequency. Taking advantage of this situation, the subcarrier is chosen at the upper end of the luminance signal spectrum and is an odd multiple of half the line scanning frequency, causing the sidebands of the subcarrier to interlace with energy bunches of the luminance signal (Fig. 2).

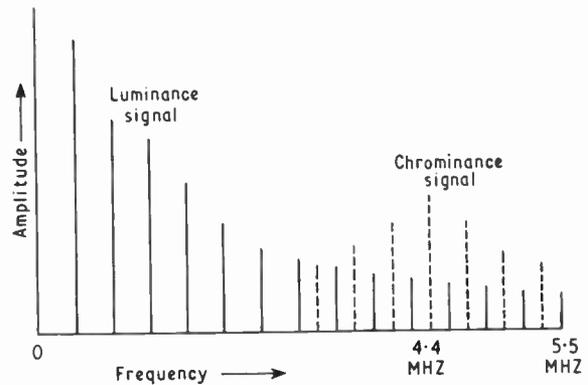


Fig. 2. Spectrum of interleaved luminance and chrominance signals.

This gives rise on the compatible monochrome picture to a pattern of dots due to the subcarrier interlacing on successive lines and fields repeating at 12.5 Hz, giving a low visibility pattern. It must be remembered that this situation is only exactly as described when the picture content is stationary, but it is maintained sufficiently in practice when movement occurs.

In addition to specifying *I* and *Q* in terms of proportions and signs of *R* - *Y* and *B* - *Y*, the actual constants or weighting factors are chosen such that the subcarrier modulated by the colour difference signals shall not give rise to excessive modulation depth when the composite signal is modulated upon the radio frequency carrier, whilst obtaining the best signal to noise performance possible.

If a receiver is to recover the colour difference signals it must be provided with a phase reference.

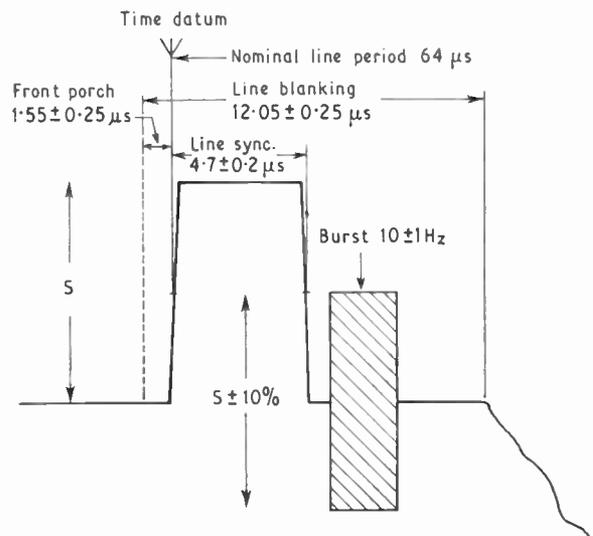


Fig. 3. Colour synchronizing waveform.

This is the colour synchronizing signal, usually referred to as the 'colour burst'. This is shown in Fig. 3 and consists of 10 cycles of subcarrier of reference phase $+180^\circ$, i.e. $-(B-Y)$. The receiver generates a continuous reference subcarrier signal of constant phase relationship to the burst by using either a phase control loop or a narrow band filter. The receiver normally uses two synchronous demodulators supplied with the modulated subcarrier and the locally-generated reference signal of appropriate phase. It is not essential to recover the original I and Q signals and it is in fact usually more convenient to recover $R-Y$ and $B-Y$, since these signals may be used in the receiver with considerably less complexity. By the use of appropriate phases of the reference signal any transformation of axes may be carried out. These axes need not be orthogonal, it being common practice in the U.S.A. to use two axes 57° apart known as the X and Z axes.

If signals other than the originally modulated I and Q signals are recovered, the bandwidth of both signals after demodulation must be restricted to that of the narrower Q signal, if quadrature crosstalk on colour transitions is to be avoided. This is because both sidebands are necessary to separate the two signal components (see Fig. 4).

3. Limitations of the N.T.S.C. System

The N.T.S.C. system is most satisfactory in many aspects, and in particular the gated nature of the burst infers an excellent noise performance. Further, the suppressed carrier modulation of the colour information means that the compatibility is good since the amplitude of the subcarrier is zero for monochrome parts of the transmitted scene and in the coloured parts it is proportional to the saturation.

The sensitivity of the N.T.S.C. system is due to the hue information being carried by the phase of the subcarrier, the eye being particularly sensitive to hue errors, especially in respect of flesh tones and other easily recognized colours. The eye can perceive the hue change resulting from a change in subcarrier phase of 2° and a change of 10° can be disturbing.

There are many points in the chain between the studio analysing device and the receiver display device where phase errors can arise. Since the subcarrier 'rides' on the luminance signal, should any part of the chain give rise to level dependent phase shifts, usually referred to as differential phase, hue errors will arise. If one of the sidebands be partially removed, for instance by receiver mistuning, short-term multi-path or misalignment, colour errors at transitions will arise due to crosstalk from one colour signal to the other (Fig. 4). Longer term multi-path reception can contaminate the reference burst with

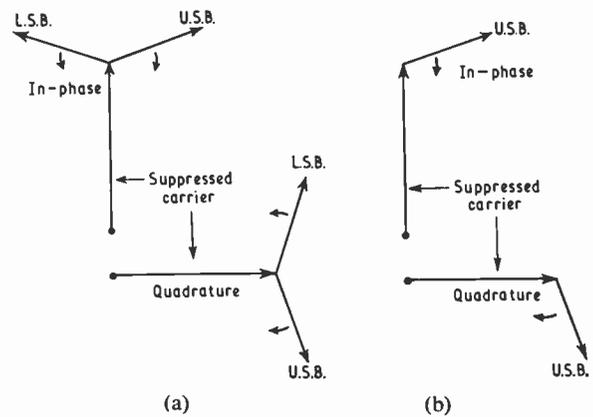


Fig. 4. Colour errors due to crosstalk.

- (a) Double sideband. No crosstalk, projections of quadrature sidebands upon in-phase axis are equal and opposite.
- (b) Single sideband. Crosstalk occurs, projections of quadrature sideband upon in-phase axis is not cancelled.

picture colour information, giving yet a further possibility of hue errors. The N.T.S.C. receiver requires a viewer control to set the phase of the reference generator output to the correct phase relationship with the colour subcarrier signal, usually called the hue or tint control and there can be no satisfactory reference to ensure accurate adjustment. Finally, although any individual phase error may be fairly small, the viewer receives the sum of all the various phase errors.

4. The PAL System

The PAL system retains all the desirable features of the N.T.S.C. system, but eliminates the sensitivity to phase changes.

In its original form the PAL system used I and Q modulation as N.T.S.C., but certain minor changes have been effected in the PAL system to simplify receiver circuits without adversely influencing performance. The two modulation axes have been shifted by -33° and are now

$$0.493(B-Y) = U \quad \text{and} \quad 0.877(R-Y) = V \quad \dots\dots(6)$$

This permits the receiver demodulators to recover signals that are suitable for application to the display tube after amplification without matrixing operations, and decoder adjustments are likewise simplified. To simplify notation the two colour difference signals, $(B-Y)$ and $(R-Y)$, including their weighting factors are now referred to as U and V respectively. Quadrature modulation as in N.T.S.C. is used, but the conjugate signal is transmitted on alternate lines. If a given line uses the signal $U+jV$, the next line in time uses $U-jV$, inferring that the V signal is reversed in sign from line to line. A switch must be added to the N.T.S.C.

receiver to remove this line by line inversion. The effect of this when there is some phase error can be seen in Fig. 5. On one line the hue is shifted from the wanted magenta towards red and on the next line in time towards blue. Thus phase errors result in colour errors that are equal and opposite; the eye tends to average the differences in colour on adjacent lines and thus no hue error should be seen.

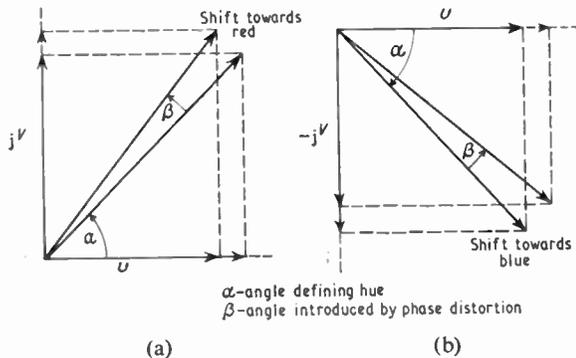


Fig. 5. Effect of phase error upon chrominance vector. (a) Line $2n-1$. (b) Line $2n$.

For several reasons, including the action of the eye itself and the non-linearity of the display device characteristic, this averaging is only approximate. When the error in phase exceeds 10 or 15°, the difference between adjacent lines becomes visible. At very close viewing distances the actual colour change from line to line is visible and at greater distances a brightness change is observed. Figure 6 shows that for an interlaced raster pairs of spatially sequential lines from an interlaced pair of rasters have the same error giving rise to a relatively coarse pattern, usually known as 'Hanover blinds'.

5. Delay Line Demodulation

If the averaging process were to be carried out electrically rather than by using the viewer's eye, the disadvantageous situation described above would not occur. In practice an acoustic delay line of effectively one scanning line delay duration is used, permitting the operations shown in Fig. 7 to be carried out. In fact, the line must be of such a length that it contains the integral number of 1/2 cycles of subcarrier nearest to the duration of one scanning line. The sum and differences of signals occurring on time sequential lines are taken:

Line $2n-1$	$U+jV$	$U+jV$
Line $2n$	$U-jV$	$U-jV$
Sum	$2U$	difference $2jV$

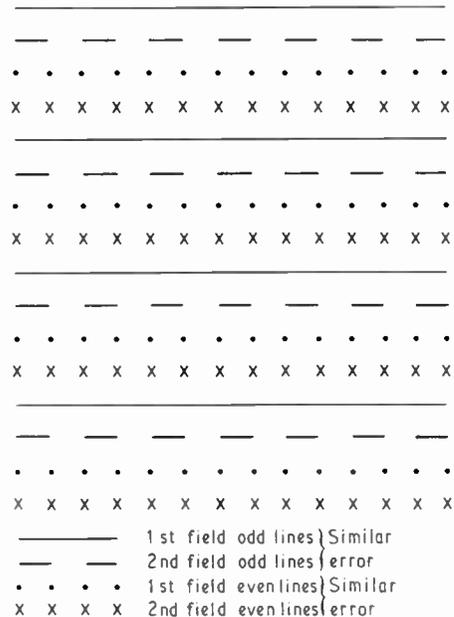


Fig. 6. Coarse structure from optical averaging of largish phase errors.

Line $2n$	$U-jV$	$U-jV$
Line $2n+1$	$U+jV$	$U+jV$
Sum	$2U$	difference $-2jV$

It can be seen that the U and V signals have been separated without reference to demodulation, the V signal still alternating in sign. This may be compared with the N.T.S.C. situation in the presence of phase errors (Fig. 5(a)), where a shift in phase must change the ratio V/U and hence the reproduced hue. Where the signals can be separated by means of the delay line as in PAL, any shift in phase can only effect the signals similarly reducing both of them after demodulation by a factor of $(1-\cos \beta)$, where β is the phase change. No change in the ratio V/U can occur, hence there is

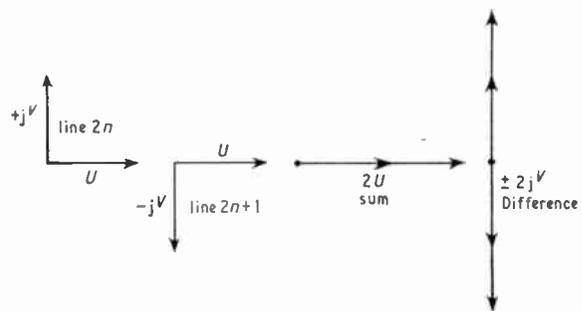


Fig. 7. Separation of U and V with delay line.

no change in hue but only a change in saturation. This action can be clearly seen if eqn. (5) is expressed in exponential form:

$$M = S e^{j\alpha} \cdot e^{j\omega_0 t} \quad \dots\dots(7)$$

where S is the saturation and α is the angle $\tan^{-1} V/U$, as in eqn. (5). The modulation of the subcarrier may now be written as:

$$M = S e^{j\alpha} \quad \dots\dots(8)$$

The sequence of lines in PAL is now:

Line $2n-1$ $2M = S e^{j\alpha}$

Line $2n$ $2M^* = S e^{-j\alpha}$

Sum $2(M+M^*) = \overline{S(e^{j\alpha} + e^{-j\alpha})} \quad \dots\dots(9)$

Difference $2(M-M^*) = S(e^{j\alpha} - e^{-j\alpha}) \quad \dots\dots(10)$

Thus the sum is

$$S(e^{j\alpha} + e^{-j\alpha}) = 2S \cos \alpha = 2U \quad \dots\dots(11)$$

and the difference is

$$S(e^{j\alpha} - e^{-j\alpha}) = 2j \sin \alpha = \pm 2jV \quad \dots\dots(12)$$

If the carrier term $e^{j\omega_0 t}$ is included, the above equations (11) and (12) become

Sum = $2U \cos \omega_0 t \quad \dots\dots(13)$

and

Difference = $\pm 2V \sin \omega_0 t \quad \dots\dots(14)$

This illustrates how the use of the delay line enables the subcarrier modulated in both amplitude and phase to be separated into two purely amplitude modulated signals.

6. Phase Errors and Reproduced Hue

The signals of equations (13) and (14) are demodulated in synchronous detectors with the appropriate phase of reference carrier injected to recover $2U$ and $2V$. If the carrier signal is shifted in phase relative to the reference by the angle β or if the injected reference phase is incorrect by the angle β the recovered signals will be $U \cos \beta$ and $V \cos \beta$. Equation (3) shows that the hue is

$$\alpha = \tan^{-1} V/U$$

Since $\tan^{-1}(V \cos \beta / U \cos \beta) = \tan^{-1} V/U$, the hue is seen to be unchanged.

A desaturation ΔS will occur:

$$\Delta S = S(1 - \cos \beta) \quad \dots\dots(15)$$

The use of the delay line eliminates changes of hue when either the signal itself suffers a phase change relative to its reference burst or if the reference generated in the receiver decoder has not the correct phase relationship with the signal to be demodulated,

Thus no viewer control is required. A 30° phase shift gives rise to a saturation change of only 10%.

7. Single-sideband Distortions

The crosstalk of one colour difference signal into the other when demodulating after the partial elimination of one sideband with the N.T.S.C. system has already been briefly considered. This distortion can also be eliminated when using the PAL system with a delay line in the decoder. To investigate the action in greater detail the U channel signal will be taken as a constant function of time and the V channel a repetitive rectangular signal of unit amplitude:

$$f_i U = 1 \quad f_i V = \frac{1}{2} + \sum a_m \cos \omega_m t \quad \dots\dots(16)$$

These are modulated for N.T.S.C. on the two subcarriers $\cos \omega_0 t$ and $\sin \omega_0 t$ respectively. The resulting signal is

$$M = -(\frac{1}{2} + \sum a_m \cos \omega_m t) \sin \omega_0 t + \cos \omega_0 t \quad \dots\dots(17)$$

If the upper sideband is eliminated the signal becomes

$$M = -[\frac{1}{4} + \sum (a_m/2) \cos \omega_m t] \sin \omega_0 t + [\frac{1}{4} + \sum (a_m/2) \sin \omega_m t] \cos \omega_0 t \quad \dots\dots(18)$$

It can be seen that there is a crosstalk component from V into U , and that the spurious components are shifted in phase by 90° . If this analysis is applied to the PAL system, line $2n-1$ will be as equations (16), (17) and (18). Line $2n$ when both sidebands are present, will be

$$M^* = +(\frac{1}{2} + a_m \cos \omega_m t) \sin \omega_0 t + \cos \omega_0 t \quad \dots\dots(19)$$

Similar band limitation gives rise to the signal

$$M^* = +[\frac{1}{4} + \sum (a_m/2) \cos \omega_m t] \sin \omega_0 t + [\frac{1}{4} - \sum (a_m/2) \sin \omega_m t] \cos \omega_0 t \quad \dots\dots(20)$$

The PAL delay line decoder takes the sum and difference of equations (18) and (20). Adding lines $2n-1$ and $2n$, all the terms containing ω_m are equal in magnitude and opposite in sign and are eliminated, leaving $\cos \omega_0 t$ free of crosstalk. Subtracting:

$$M - M^* = -(\frac{1}{2} + a_m \cos \omega_m t) \sin \omega_0 t + (a_m \sin \omega_m t) \cos \omega_0 t \quad \dots\dots(21)$$

The first term is the original signal free from crosstalk and the second term will not be demodulated provided the phase of the reference subcarrier applied to the detector is correct. Although these signals are now free of crosstalk, they are at half amplitude since the delay line normally gives outputs of $2U$ and $\pm 2V$. Thus, as in the case of phase distortion the hues are correct but the saturation is reduced by one half where only one sideband of the transient is present.

8. PAL with Simultaneous Phase and Single-sideband Errors

The analysis above shows that if phase errors are added to a signal where there has been partial elimination of one sideband, the non-alternated axis will still be reproduced clear of errors. The saturation will be reduced by the factor of unity minus the cosine of the error angle and by 50% over the part of the spectrum where the sideband has been eliminated.

The result of a similar situation in the alternated axis is rather different. The wanted component will still be reduced by the factor of unity minus the cosine of the error angle and 50% where one sideband is missing, but in addition a quadrature component from the other signal alternating in sign, line by line, will be introduced and its magnitude will be proportional to the sine of the error angle.

If the situation had been analysed with the modulation of *U* and *V* interchanged, the results would also be interchanged, i.e. the alternated axis would be clear of errors and the non-alternated axis would have the quadrature component. Should the two axes carry varying modulation, then a quadrature component will occur in both. The result will be as described above, all demodulated components being reduced by the factor of unity minus the cosine of the error angle and a further 50% where only one sideband is present; additionally the quadrature component will be demodulated, alternating in sign from line to line and its magnitude proportional to the sine of the error angle.

The sidebands of the modulated *U* and *V* signals will both be inherently slightly asymmetric since their

bandwidth exceeds the double sideband capability of a standard *I* channel having a 6 MHz sound-vision carrier spacing. In practice significant crosstalk does not result from this asymmetry.

9. Colour Switching Synchronization

Two types of colour synchronization are necessary in the PAL system. The first is similar to that required by the N.T.S.C. system to generate a continuous reference phase subcarrier signal, the second is to ensure that the receiver switch in the *V* channel is in step with the transmission switching. These two processes are carried out by means of the same synchronizing signal. It has been shown that the reference burst of the N.T.S.C. signal has a much better signal to noise performance than can be used. The burst in the PAL system is as that of N.T.S.C., but it is alternated in phase with the *V* signal alternation by $\pm 45^\circ$. This phase alternation reduces its effective amplitude to 0.7 of that of the N.T.S.C. burst, but the performance still has plenty in hand. The phase modulation can also be extracted from the reference generator as a half line-frequency signal and used to control the receiver switch. This synchronism is so effective that it is normally better than the locking ability of the field scanning circuit under poor signal conditions. Figure 8 shows the block diagram of a PAL decoder.

10. Chrominance Signal Lock

In the decoding process already described, a reduction of saturation with phase distortion and the re-introduction of an alternating quadrature component by phase error occurs. This effect can be reduced if the reference signal can be caused to follow

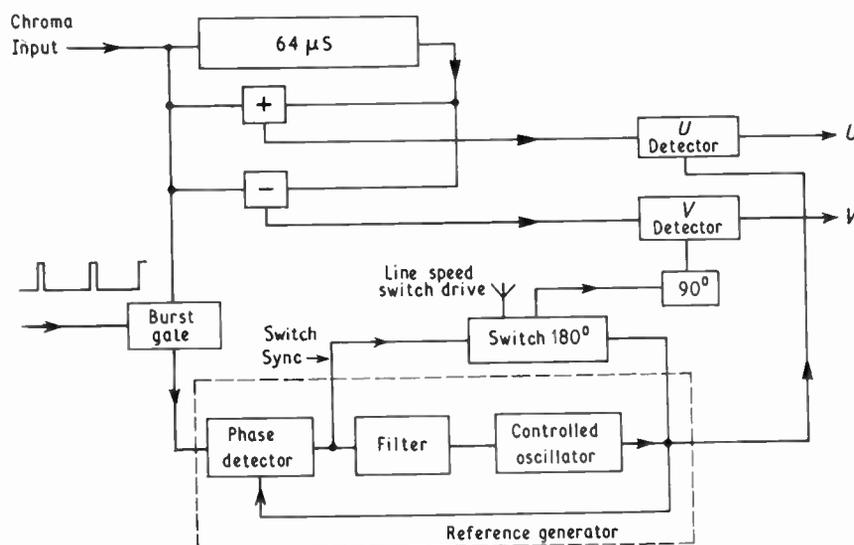


Fig. 8. Block schematic of standard PAL decoder.

the spurious phase changes. Since the delay line generates two amplitude-modulated signals clear of phase modulation from the chrominance signal, these two need only to be combined correctly to provide a reference signal since they will always be present when the scene has any colour content. The stages in the combination process are as follows:

The U and V carrier signals have a 90° phase relation, so one must be shifted. Both U and V signals, in common with all colour difference signals, can be either positive or negative sign, giving a 180° ambiguity. If the frequency of both is doubled, the 180° ambiguity becomes 360° and is thus removed, and the two signals may now be added. In practice the 90° shift is not required since with frequency doubling it becomes 180° , this being eliminated by phase inversion. Any phase changes occurring in the chrominance signal will be passed on the separated U and V signals and thence to the combined frequency doubled signal. This may be used to lock directly a low- Q oscillator which will then follow these phase variations.

Since the oscillator is locked by a double frequency signal it can have a 180° ambiguity with respect to its relation to the subcarrier reference phase. This is removed by additionally injecting the separated colour reference burst into the oscillator. While the oscillator has to be easily synchronized, it is also a requirement that the phase, once defined by the injected burst shall not deviate by more than 90° during the period of one scanning line. This is necessary since there might only be colour on the extreme right of the picture area.

The above two requirements are in no way conflicting and a suitable oscillator has been developed by G. Mahler.⁵

The oscillator can be designed to have a chosen locking bandwidth. If it is too narrow, rapid phase changes will not be followed; should it be too wide the output phase will be unduly influenced by noise under poor reception conditions. Typically the effective bandwidth is approximately 100 kHz (see Fig. 9).

The desaturation caused by phase errors and the results of the combination of phase errors with single-sideband distortion are eliminated since the reference subcarrier generated in the decoder always has the correct relationship to the chrominance signal. Further, split screen techniques can be used without equalizing the phase of the two encoders to a high degree of accuracy.

Another version of this decoding method avoids the requirement for the switch in the reference subcarrier supply to the V demodulator. A separate locked oscillator is used for the U and V demodulator

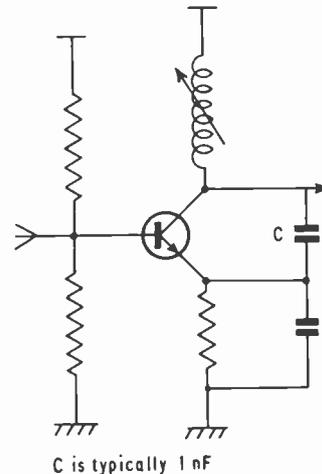


Fig. 9. Injection locked oscillator.

reference subcarrier. The burst is also resolved into two components, a constant one in the U axis and one alternating by 180° in the V axis. These are used to set the initial phase on the locked oscillators, the alternating burst applied to the V channel oscillator forces it to change in phase by 180° from line to line, eliminating the need for the switch and its associated synchronizing circuit.

Owing to their additional complexity these types of decoding circuit have been restricted to use in certain types of studio equipment. (They are often referred to as 'New PAL' and 'New new PAL', this kind of terminology is to be deprecated.)

11. PAL Subcarrier Frequency Offset

The N.T.S.C. subcarrier frequency gives a low-visibility dot pattern on the compatible monochrome receiver with a half-line offset relationship to the scanning frequency:

$$\text{line scanning frequency} = f_{sc} \times 2/567$$

the complete pattern having a repetition frequency of 12.5 Hz. When PAL is used the 180° line by line alternation of the V axis effectively lines up the subcarrier dots on adjacent lines, transforming the interlaced pattern into vertical rows of high visibility. If a $\frac{1}{4}$ -line offset is used, both switched and non-switched axis signals will interlace, but the angle joining the dots on one line to those on the line below will be half that of the $\frac{1}{2}$ line offset condition. The angle will be positive for non-switched and negative for switched axis components. A further improvement in compatibility is possible if the pattern interlaces from field to field and this is achieved by offsetting the subcarrier frequency by half the field scanning frequency in addition to the $\frac{1}{4}$ line offset. This, in fact produces a dot pattern that repeats over eight

fields. It has been shown recently that, for systems with certain numbers of lines, additional half-field offset is not required. If h is the number of lines used by the system and if $(h+1)/8$ or $(h+7)/8$ is an integer the half-field offset is required, but not if $(h+3)/8$ or $(h+5)/8$ is an integer. Hence the 625-line system requires the additional offset, but not 525, 405 or 819 lines. The complete line frequency/subcarrier relationship for the 625-line system is:

$$\text{line scanning frequency} = \frac{f_{sc} - \frac{1}{2}f_{\text{field}}}{284 - \frac{1}{4}}$$

If the line scan frequency is taken as 15625 Hz, then the now agreed subcarrier frequency for PAL of 4.43361875 MHz \pm 1 Hz is obtained.

12. Comb Filter Action of the Delay Line

Since the luminance and chrominance signals share the upper part of the video band, interference between them can occur. Where high frequency luminance components associated with transients are demodulated by the chrominance demodulators, the effect is described as crosscolour. It has the appearance of strings of coloured dots at the luminance transient. The exact appearance of the crosscolour is defined by the relationship between subcarrier and line scanning frequency and the angle the transient makes with the scanning axes.

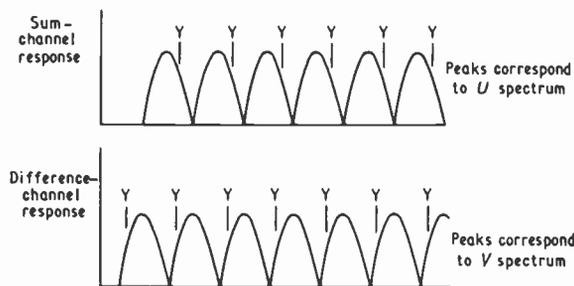


Fig. 10. Delay line separation of U , V and Y spectra.

In the PAL decoder using a delay line an additional action takes place. If a continuously changing frequency is applied to the delay line including its summing circuit, the output will be large when the delayed and undelayed signals correspond in phase and will be small when a 180° phase difference occurs. There will thus be a succession of peaks and troughs in the amplitude frequency response, their separation will be $1/T$ where T is the delay time, i.e. a peak and trough for each multiple of the line scanning frequency. In fact, when the precise delay is set correctly for PAL the peaks in the sum and difference channels correspond to the U and V spectra respectively and the

luminance spectral components lie as shown in Fig. 10. A significant reduction in the level of the luminance components in the chrominance channel is achieved and a consequent reduction in crosscolour by 3 dB. Yet a further reduction in crosscolour takes place in the PAL decoder since the luminance components in the V channel are alternated in sign from line to line and their visibility reduced.

13. The Delay Line

If operation is to be satisfactory the delay time must be precisely equivalent to 284 cycles of the subcarrier when the standard frequencies are used for colour subcarrier and horizontal scanning frequency. The accuracy has to be within 3 ns and until recently it was necessary to provide a means of trimming the delay time to the exact value by associating an adjustable electrical line with the acoustic line. These lines are in the form of an encapsulated glass rod with a piezo-electric transducer, at each end, exciting a shear mode vibration. Recently lines have become available that can be manufactured to the required tolerance. These use a V-shaped path having total reflection, and the delay time can be adjusted by machining the surface where the reflection takes place after the transducers have been attached. A further convenient feature of these lines is that the impedance is about 150Ω rather than the low value of 50Ω for the rod lines.

14. PAL Signal Error Correctors

It is possible to remove errors, both of phase and amplitude, at any point in the transmission chain rather than to carry out the process once and for all at the receiver decoder. This process may be carried out by demodulation using the chrominance-lock technique and re-modulation using a stable subcarrier obtained from the incoming signal burst.

There is an alternative method of correction which does not involve demodulating the chrominance signal, but makes use of an added correction signal. If the phase distortion is considered as a phase shift of the chrominance modulation phasor, its effect can be removed by adding to it a phasor 90° removed from the wanted signal in phase and of appropriate magnitude. Figure 11 shows that the required correcting signal is $-jS \sin \beta$, the derivation of which is as follows.

If the chrominance signal $S.e^{j\alpha}.e^{j\omega_0 t}$ modulates a carrier of double frequency $e^{j2\omega_0 t}$ in a balanced modulator the output will contain two components, one of the same frequency spectrum as the input signal and one of triple frequency. Only the first of these is of interest, the other being filtered out. When the phase of the double frequency carrier is suitable the relationship between the input and the

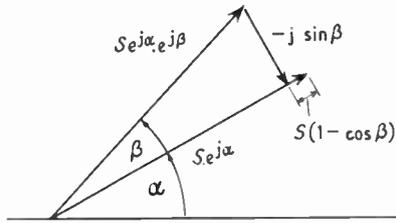


Fig. 11. Error correction by addition of compensating signal.

wanted output signals is as follows. The action of the modulator is to multiply the input signals:

$$(S \cdot e^{j\alpha} \cdot e^{j\omega_0 t}) \cdot (e^{j2\omega_0 t}) = (S \cdot e^{j\alpha} \cdot e^{j3\omega_0 t}) + (S \cdot e^{-j\alpha} \cdot e^{j\omega_0 t}) \quad \dots\dots(22)$$

The first resultant is the rejected triple frequency component, the second is the wanted output, but transformed to the conjugate signal. The output of the modifier is thus similar to the input signal, but with the opposite alternation sequence.

If the signal has an added phase error this will also be transformed, i.e. if the error angle is β and the distorted signal is $S \cdot e^{j\alpha} \cdot e^{j\beta}$, the modified signal will be $S \cdot e^{-j\alpha} \cdot e^{-j\beta}$. On the following line the signals would be $S \cdot e^{-j\alpha} \cdot e^{j\beta}$ at the input and $S \cdot e^{j\alpha} \cdot e^{-j\beta}$ modified.

The modified signal is delayed for the period of one line, using a normal PAL delay line and the difference taken between the input and the delayed, modified signals, this giving the required correction signal:

$$S \frac{e^{-j\beta} - e^{j\beta}}{2} = -jS \sin \beta \quad \dots\dots(23)$$

The action of adding the correcting signal can be clearly seen if the right-hand side of eqn. (23) is combined with the phase error modulation component:

$$(S \cdot e^{j\beta}) + \left(S \frac{e^{-j\beta} - e^{j\beta}}{2} \right) = S \left(e^{j\beta} + \frac{e^{-j\beta} - e^{j\beta}}{2} \right) \quad \dots\dots(24)$$

$$= S \frac{e^{j\beta} + e^{-j\beta}}{2} = S \cos \beta \quad \dots\dots(25)$$

Equation (25) shows that the phase error $e^{j\beta}$ has been converted to a change in saturation from S to $S \cos \beta$.

In such an error corrector it is desirable to restore the saturation error introduced by the phase correction. It is similarly possible to derive a suitable correction signal, provided a reasonable maximum phase error is prescribed above which correction is not required.

From eqn. (25) the correction signal is seen to be

$$S(1 - \cos \beta) = 2S \sin^2 \beta/2 \quad \dots\dots(26)$$

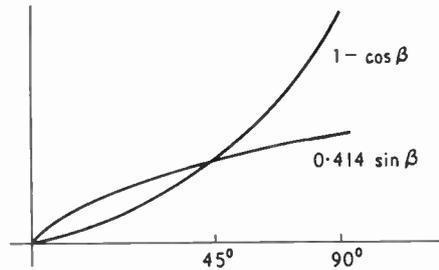


Fig. 12. Comparison between amplitude correction using $(1 - \cos \beta)$ and approximation by $K \cdot \sin \beta$.

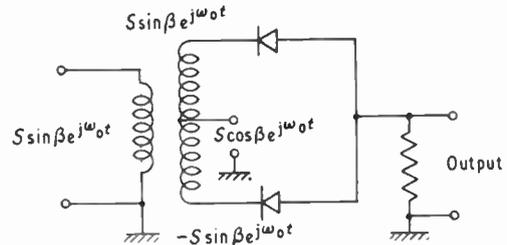


Fig. 13. Automatic commutating switch for amplitude error corrector.

If β is taken as positive and not considered above the chosen maximum, eqn. (26) may be approximated by

$$S(1 - \cos \beta) = S \cdot k \cdot \sin \beta \quad \dots\dots(27)$$

The required signal is obtained by shifting the phase correcting signal of eqn. (23) by 90° and multiplying by the factor k . Figure 12 shows the approximation when the upper limit of correction corresponds to a phase error of 45° . It can be seen that below the selected point the correction is slightly too great and beyond this point the correction falls off.

This solution only applies for positive error angles, for a negative error angle the correcting signal has the wrong polarity. It is necessary to incorporate means of commutating the saturation correcting signal when the error phase is negative with respect to the phase reference axis. A phase-splitting transformer will provide the two polarities of the correcting signal $S \sin \beta$ and $-S \sin \beta$, which may be selected by means of a diode switch. A controlling signal is required to operate the switch when the error angle β changes from negative to positive with respect to the zero phase axis. A suitable signal is $S \cos \beta$ since, although $\sin \beta$ changes in sign with the change of the angle β , $\cos \beta$ does not change.

The signal $S \cos \beta$ can be obtained in a manner analogous to the phase correcting signal, but by adding the input signal and the delayed modified signal instead of taking the difference:

$$S \frac{e^{-j\beta} + e^{j\beta}}{2} = S \cos \beta \quad \dots\dots(28)$$

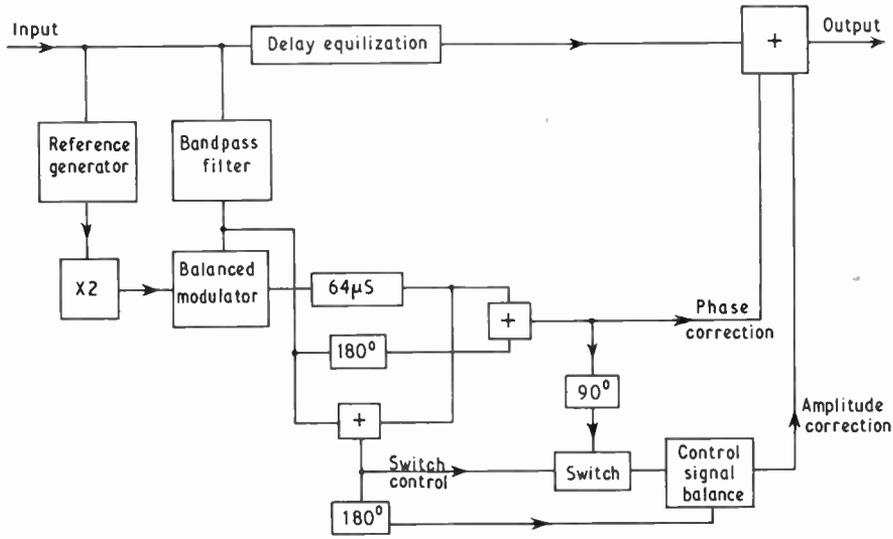


Fig. 14. Block schematic of PAL error corrector.

The action is illustrated in Fig. 13. Each diode has one of the phases of the correcting signal applied to it, either $S \sin \beta$ or $-S \sin \beta$. Both diodes also receive the control signal $S \cos \beta$. One diode will receive $(S \cos \beta + S \sin \beta)$ and the other $(S \cos \beta - S \sin \beta)$. For all values of β other than the trivial ones of 0° and 90° the sum will be greater than the difference for positive values of β and the difference will be greater for negative values of β . The circuit is so arranged that the diode with the greater voltage developed across it shall conduct, commutating the correction voltage with the change in sign of the error angle with respect to the zero phase axis. A complete practical circuit would require two sets of switching diodes, since as described only half cycles of the correcting voltage would be supplied. Additionally, provision must be made for cancelling the control signal that would otherwise appear in the output. This type of automatic commutator is the result of a suggestion by G. Mahler.⁶ A block diagram of a complete corrector is shown in Fig. 14.

15. Effects of the Averaging Process

The PAL system depends upon averaging the colour information carried by sequential lines to eliminate errors. Where a rapid vertical chrominance transition occurs this will be influenced by the averaging operation, in fact, the effects are difficult to show using even critical picture content. More than one averaging process can occur, as when an error corrector is followed by a decoder using a delay line.

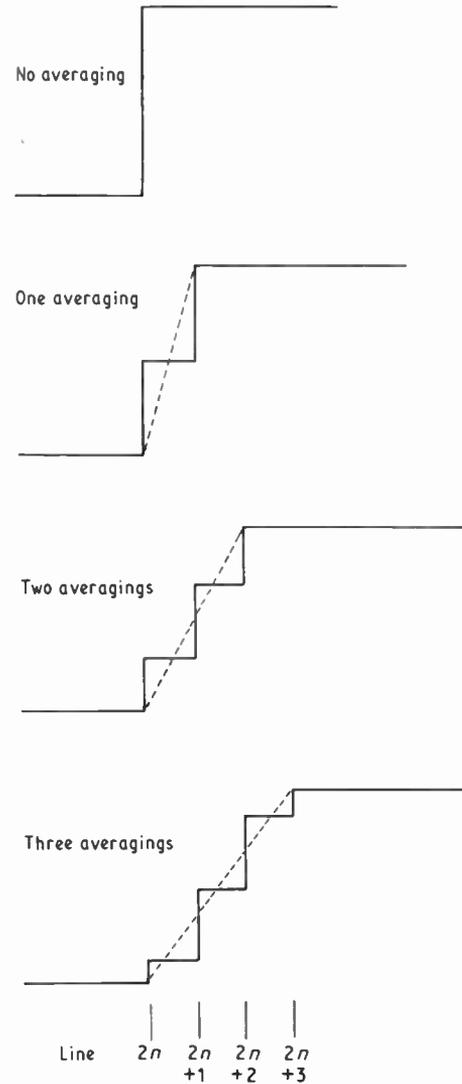


Fig. 15. The effect of multiple averagings upon a colour difference transition.

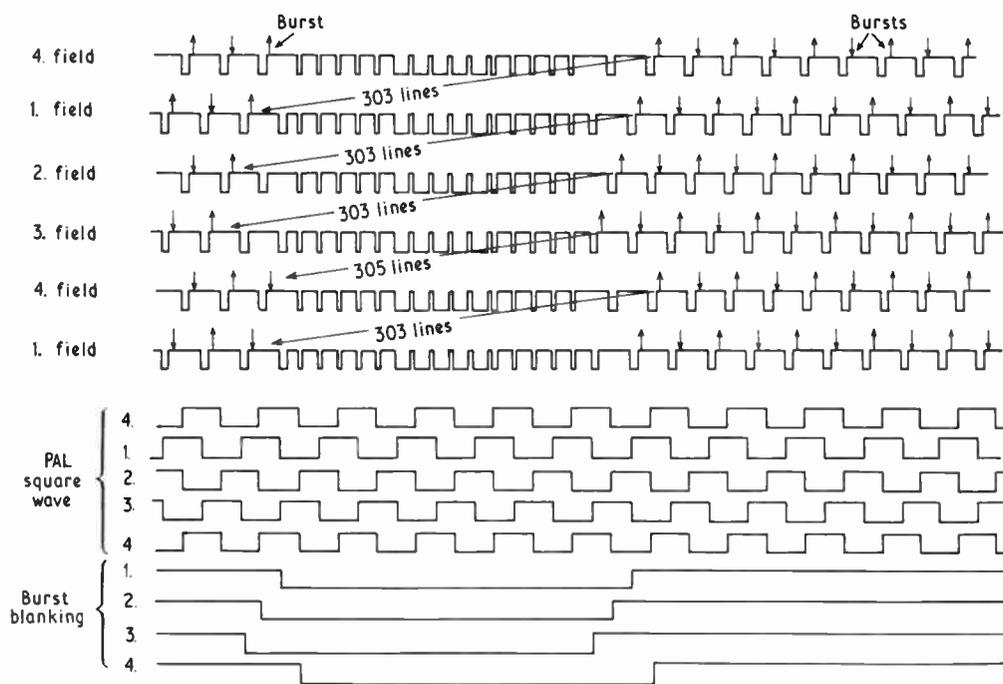


Fig. 16. Burst suppression sequence.

Two effects are of importance, the transition is slowed down and the centre point of the transition is delayed. Figure 15 shows the effects of from zero to three averaging processes, i.e. from simple non-delay line decoding to the situation arising from two error correctors followed by delay line decoding. The centre of the transition is moved downwards, i.e. delayed, by the equivalent of half the height of one scanning line by each averaging. When only one averaging is involved the effects are of no consequence, but with two averagings as when an error corrector is introduced, the luminance signal can with advantage be delayed by one line, matching the downward shift of the centre of the colour transition. The equalization is very effective since lack of time coincidence between luminance and chrominance is much more disturbing than the change in resolution.

16. Burst Blanking Sequence

The N.T.S.C. system suppresses the reference burst during the field synchronizing interval. If the burst is similarly suppressed when using the PAL system, the phases of bursts starting and ending fields will execute a four-field sequence, pairs of fields starting and ending with negative and positive phase bursts (Fig. 16). Normally this causes no difficulties, except when certain types of reference generators are detuned near to the limit of operation. The starting phase after the field synchronizing interval may differ for the two

starting and ending burst conditions, giving rise to a slight flicker at the top of the picture when high saturations occur.

The flicker effect described is removed by adopting the special burst blanking sequence shown in Fig. 16 so that all fields start and end with the same burst phase. The use of this special burst blanking gives rise to a problem when video tape recording the signal, since the recording machine uses the bursts in the phase stabilization circuits and certain bursts are missing. Although this problem has been solved, the final decision concerning the inclusion of the special blanking has been deferred pending the report of Sub-group 6 of the E.B.U. *Ad-hoc Group* on Colour Television (Studio and recording equipment). In fact, experience shows that when the N.T.S.C. type burst blanking is used, the effect on most receivers, even when extremely maladjusted, is very much of a second order.

Associated with the above discussion is a more general point. Any colour system with a line sequential process will normally involve a four-field repeating sequence. Figure 17 illustrates this point, showing that any line of a pair of interlaced fields will alternately have both of the colour sequence signals associated with it. This must be considered with regard to its effect upon means taken to bring independent signal sources into synchronism and to the run-up of recording machines.

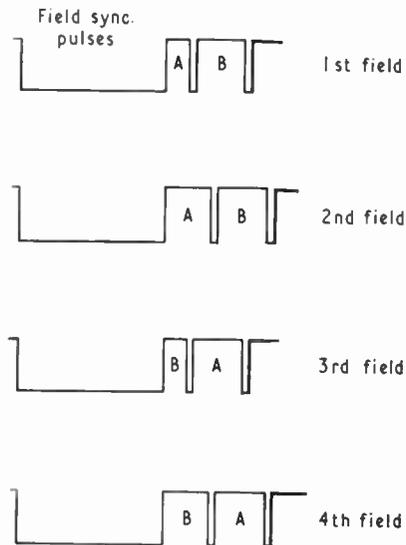


Fig. 17. Four-field pattern of a line sequential system.

17. Conclusion and Acknowledgments

The PAL system has been described in terms of its relationship to the N.T.S.C. system. The properties of the system under conditions of single and multiple distortions have been analysed and other aspects of the system have been outlined.

The author wishes to express his indebtedness to Dr. Walter Bruch of the Telefunken Company for his help and encouragement over a number of years and to thank the directors of Rank Bush Murphy Limited for permission to publish this paper.

18. References and Bibliography

1. Reports of the E.B.U. *Ad-hoc* Group on Colour Television, 1st edition—October 1963, 2nd edition—February 1965 and addendum to the 2nd edition.
2. W. F. Bailey, 'The constant luminance principle in N.T.S.C. colour television', *Proc. Inst. Radio Engrs*, **42**, pp. 60-66, January 1954.
3. D. C. Livingston, 'Colorimetric properties of gamma-corrected color television systems', *Convention Record of I.R.E.*, **1**, Part 4, pp. 51-56, 1953.
4. W. Bruch, 'PAL—Selected Papers', March 1965. 'PAL—Selected Papers', June 1966. (Both published by *Telefunken-Zeitung*, Telefunken A.G., Berlin.)
5. G. Mahler, 'Der Mitnahmeoszillator', *Telefunken-Zeitung* **37**, No. 2, pp. 136-39, 1964.
6. W. Bruch, 'Ein neues Verfahren zur zentral Phasenfehlerkorrektur eines PAL-Farbfernsehsignals', *Rundfunktechn. Mitt.*, **9**, No. 4, pp. 189-98, 1965.
7. 'Characteristics of Monochrome Television Systems', C.C.I.R. Report 308 (C.C.I.R., Geneva, 1963).

19. Appendix: Specification of the PAL System for B.B.C. 625-line Transmissions

1. *General Specification*

Luminance component
Chrominance component

Amplitude modulation of the picture carrier.
Simultaneous pair of components transmitted as amplitude modulated sidebands of a pair of suppressed sub-carriers in quadrature having a common frequency.

2. *Colour Subcarrier, f_{sc}*

$$f_{sc} = 4.433\ 618\ 75\ \text{MHz} \pm 1\ \text{Hz}$$

3. *Frequency Spectrum of Composite Picture and Sound Signals*

Vision to sound spacing
Main sideband
Vestigial sideband
Chrominance sidebands

6 MHz.
5.5 MHz.
1.25 MHz.
 $f_{sc} + 1.07\ \text{MHz (max)}$ and $-1.6\ \text{MHz (max)}$.

4. *Transmitted Colour Picture Signal Waveform*

Colour synchronization

Complies with C.C.I.R. monochrome television specifications⁷ with following modifications:

Subcarrier burst

Duration: $10\ \text{Hz} \pm 1\ \text{Hz}$

Start: $5.5 \pm 0.2\ \mu\text{s}$ after line sync datum

Amplitude: 0.5 ± 0.1 of line sync amplitude.

Omitted during field blanking periods for 9 lines as in Fig. 16. The phase is 135° on lines where the polarity of E'_V is positive and 225° on line where the polarity of E'_V is negative.

5. *Delay Characteristic of the Transmitted Signal* Uniform to 4 MHz
 -0.08 μs at f_{sc}
 -0.27 μs at 5.5 MHz
6. *Luminance Component*
 Attenuation frequency characteristic Uniform to 5.5 MHz, inclusion of notch-filler possible at f_{sc} .
7. *Scanning*
 Line scanning frequency $15\,625\text{ Hz} = \frac{(f_{sc} - \frac{1}{2}f_{field})}{(284 - \frac{1}{4})}$
8. *Synchronizing and Blanking Waveforms* Complies with C.C.I.R. monochrome television specification.⁷
9. *Equation of Complete Colour Signal*
 E'_Y = voltage of luminance component of the composite signal.
 E'_R, E'_G, E'_B = gamma corrected voltages corresponding to the red, green and blue signals. $E'_Y = 0.3E'_R + 0.59E'_G + 0.11E'_B$.
- $\omega = 2\pi f_{sc}$
 Phase reference + (B - Y) axis.
- C.I.E. primary colour chromaticities of picture tube for which $E'_R, E'_G,$ and E'_B (above) are suitable.
 Red ($x = 0.67, y = 0.33$)
 Green ($x = 0.21, y = 0.71$)
 Blue ($x = 0.14, y = 0.08$)
- C.I.E. white Illuminant C ($x = 0.310, y = 0.316$)
- Equivalent bandwidths of colour difference signals before modulation, $E'_B - E'_Y$ and $E'_R - E'_Y$.
 { at 1 MHz less than 2 dB down
 { at 4.0 MHz at least 20 dB down.
- Maximum time difference of transmission of component signals ± 40 ns.
- E_M = total video voltage applied to modulator of transmitter.
 $E_M = E'_Y + E'_U \sin \omega t \pm E'_V \cos \omega t$
 $E'_U = 0.493 (E'_B - E'_Y)$
 $E'_V = 0.877 (E'_R - E'_Y)$
- The polarity of $E'_V \cos \omega t$ is positive during odd lines of the first- and second-fields and during even lines of the third- and fourth-fields.

Manuscript first received by the Institution on 13th October 1966 and in final form on 5th January 1967. (Paper No. 1101/T37.)

© The Institution of Electronic and Radio Engineers, 1967

The International Federation for Information Processing

The 1968 IFIP Congress in Scotland

Her Majesty Queen Elizabeth II has graciously agreed to be Patron of IFIP Congress 68, which more than 4,000 computer experts from 40 countries are expected to attend. The Congress will be held in Edinburgh, from 5th to 10th August, 1968.

Every three years the International Federation for Information Processing (IFIP) sponsors a world Congress in one of its member countries. Prior to the formation of IFIP, an inaugural conference was held—under the sponsorship of UNESCO—in Paris (1959). Subsequent IFIP Congresses took place in Munich (1962) and New York (1965).

During the 1968 Congress approximately 250 technical papers covering all aspects of data processing from the state of the art to future developments envisaged in the information processing sciences will be delivered by speakers from many countries. The subjects under discussion will be broadly divided into mathematics, computer languages, computer equipment, scientific applications, education and business. Proceedings, containing the papers of the general scientific programme will be published after the Congress.

In addition to invited papers reviewing various aspects of computing, a large part of the programme will be devoted to submitted papers. These will cover computer applications, hardware, software and mathematics.

Subjects are broadly divided into the following categories:

Computer applications

Applications in physical and life sciences. Applications in engineering. Applications in linguistics. Artificial intelligence. Applications in library science. Applications in management and business. Applications in social sciences. Applications in the arts and the humanities. Education.

Hardware

Analogue and hybrid computers. Computer systems. Real-time systems. Components and circuits. Graphical display and input. Data transmission.

Software

Operating systems. Programming languages. Compilers and other language processors. Parallel programming. Data structures.

Mathematics

Computational methods in analysis. Computational methods in algebra. Combinatorial and discrete mathematics. Theory of machines. Theory of algorithms.

An appeal for papers has been made and full details of the procedure for submitting papers can be obtained from IFIP Congress Office, 23 Dorset Square, London, N.W.1. Synopses must be submitted before 30th November 1967.

IFIP Congress 68 is organized by a committee formed by the British Computer Society and representing a wide cross-section of professional societies and people concerned with computers and data processing. The committee has members from the British Computer Society, the Scottish Office, the Ministry of Technology and twelve professional societies of the United Kingdom Automation Council, including the I.E.R.E. whose representative on the organizing committee is Mr. A. St. Johnston (Member).

Concurrently with the Congress an exhibition will be held at the Waverley Market in Edinburgh. Here a large exhibition area of some 30 000 ft² (2800 m²) will be available for a comprehensive display of computers, computing equipment and allied services. Digital and analogue computers, data collection-systems and data transmission facilities, logic modules, sub-assemblies, etc., will be on show in addition to working data processing installations. Between 70 and 100 leading computer manufacturers and marketing organizations, service companies and major users from all parts of the world are expected to participate in this unique international event.

Expansion of IFIP Activities

The eleventh general assembly of IFIP, recently held in Jerusalem, decided to modify the Federation's structure to encourage the formation of special interest groups on an international basis. The first of these will be the administrative data processing group, and its aims are to promote and co-ordinate research, education, and the exchange of experience in the field of information processing as applied to organizational, economic and administrative problems in public and business administration.

On the technical side of their work, the general assembly approved the formation of a new working group (WG.3.1) to conduct research into the teaching of information processing in secondary schools. In addition the formation of a new committee (TC.4) is proposed which will study techniques and applications in medical research and biomedical practice, with particular reference to the international codification of medical information.

The growing influence of IFIP was reflected in the admission of Hungary, and the forthcoming admission of Yugoslavia which will bring the total membership to 27 nations.

Cryotron Storage Cells for Random Access Memories

By

V. L. NEWHOUSE, Ph.D.†

AND

H. H. EDWARDS‡

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E. Conference on 'Applications of Thin Films in Electronic Engineering' held in London on 11th-14th July 1966.

Summary: Cryotron storage cells are attractive for random access memory because of their broad tolerances. In the past, their lower size limit has been set by the fact that their read-out signal becomes too small and short to be sensed by conventional amplifiers. A new sensing method has been developed which removes this limitation, so that the lower limit of cell size is now determined by fabrication technology only.

This paper describes this technique as well as a number of old and new cryotron cells compatible with it. Cell dimensions are calculated in terms of fabrication parameters and a comparison of the densities of various bit and word organized designs is made.

1. Introduction

Considerable effort has been devoted to the development of a practical superconducting film memory. The high density and simple structure of the continuous film memory (c.f.m.) make it a favourite contender. However, the c.f.m. has been found to have uncomfortably narrow tolerances which are strongly dependent on material parameters.¹ Thus

interest has revived in cryotron memories in which tolerances can be broadened by adjusting the geometry of the cell. This paper presents some old and a number of new cryotron cell designs, and calculates the maximum density of each. The results are summarized in Table 1. It is shown that with present constructional techniques, bit-organized designs give densities of around 10^3 cells/in², while word-organized designs can give densities of over 10^4 cells/in².

Table 1

Cell No.	Figure			Insulated cross-overs	T junctions	Pb-gate contacts	Density shown in Fig. (cell/in ²)	Proposed density† (cell/in ²)	Organization
1	3, 4	$3w_g + 6r + 3m$	$4w_c + 6m + 3r_1 + c$	9	6	3	1 150	2 300	Word
2	6, 7	$w_g + 5w_c + 2r + 4m$	$w_c + 4m + 6r_1 + 4c + L$	3	4	4	1 540	2 100	Bit
3	8, 10	$w_g + 3r + 3m + c + \alpha w_c$	$w_c + 4m + 4r_1 + 2c + w_g$	5	4	3	1 200	2 600	Bit
4	11	$w_g' + \beta w_g' + 2m + b$	$2w_c + \gamma w_c + m + 2r$	2	2	0	10 000	10 000	Word
5	12	$w_g' + 2r + m$	$w_c + 2r_1 + c$	2	2	1	28 600	36 000	Word

† Density with the following values of fabrication parameters (dimensions given in inches $\times 10^{-3}$):

Parameters	Definitions	Parameters	Definitions
$w_g = 6$	gate width of c.f.c. having current gain.	$L = 5$	length of in-line cryotron.
$w_c = 1$	control width.	$\alpha = 4$	ratio of widths of non-active write line to control width in Fig. 10.
$r = 0.5$	minimum overlap of one etched metal film over another.	$\beta = 2$	ratio of widths of gate film under an inactive control to that under an active control of Fig. 11.
$r_1 = 2$	minimum overlap with insulation film over bottom metal film.	$w_g' = 2$	gate width of c.f.c. not requiring current gain.
$m = 1$	minimum spacing between two isolated coplanar metal films.	$\gamma = 4$	ratio of widths of the inactive control to the active control in Fig. 1.
$c = 2$	minimum contact length needed to establish superconducting contact between two superimposed films.	$b = 4.5$	length of inductive leg of Fig. 11 required to give sufficient output signal.

† Research and Development Center, General Electric Company, Schenectady, New York.

‡ Formerly with Research and Development Center, G.E.C., New York; now at Biology Department, Rensselaer Institute of Technology, U.S.A.

A simple cryotron storage cell² (see Fig. 1(a)) is sensed by destroying the current stored in it, and observing the polarity of the resulting voltage pulse on the digit line. When this cell is made as small as present techniques allow, the sense signal amplitude and duration are too small to be detected by conventional methods. A new method has been conceived which makes it possible to *stretch cell currents in time*, so that they can be sensed by a cryogenic amplifier. Following a brief description of the new sensing method, this paper deals with a number of cryotron cells compatible with it, and calculates their cell dimensions and time constants in terms of fabrication parameters.

2. Current Stretching

The use of the current stretching principle will be described in connection with Fig. 1(a). This figure shows a portion of a word-organized memory. The '1's and '0's are stored in the form of clockwise and anti-clockwise currents in the cryotron cells. To read a specific word, current is passed through the word line, thus making its cryotrons resistive. A voltage is thus induced across each of these cryotrons while the current circulating in the cryotron cells is decaying.

To estimate the magnitude of this voltage signal, consider Fig. 11, which shows possible dimensions for a realization of the circuit of Fig. 1. Assuming a thickness of 5000 Å for both the insulating and tin films and a tin film resistivity of 0.2 μΩ-cm, a cell time constant for both reading and writing of approximately 5 ns results. The quenching of a 100 mA stored current would produce a read-out signal with

a peak value of approximately 150 μV. To achieve this output voltage, the rise-time of the word line read-out pulse must be considerably less than 5 ns. Clearly, this voltage is too small and of too short duration to be sensed by conventional methods.

The current-stretching technique is illustrated in Fig. 1(c).

A superconducting current $I_c(0)$ is stored in the small loop of the figure. If a resistance r is introduced into the loop, a 'read-out' voltage appears. Although the current through r will decay to zero, the voltage produced during the decay process causes a persistent current, I_f , to build up in the loop consisting of L_f and L_c . Since the sum of the voltages around the loop must be zero:

$$L_c \frac{dI_c}{dt} + L_f \frac{dI_f}{dt} = 0.$$

Integrating over the time I_f builds up from zero to its final value $I_f(t)$,

$$L_c [I_c(t) - I_c(0)] + L_f [I_f(t) - 0] = 0,$$

since $I_c(t) = I_f(t)$, we finally obtain

$$I_f(t) = \frac{L_c}{L_c + L_f} I_c(0).$$

Current $I_f(t)$ will persist until it is deliberately quenched. Thus the temporary voltage signal across r has been 'stretched in time' by any desired amount.

Returning to Fig. 1(a), each digit line is shunted by an amplifier with a superconducting (i.e. purely inductive) input. Thus the bit-line voltage produced during cell read-out induces a persistent current in

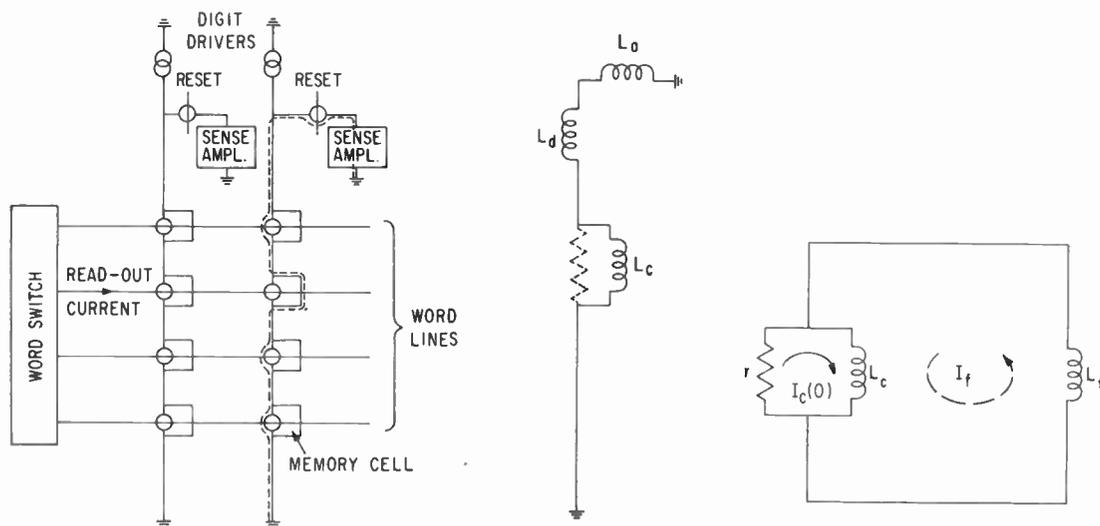


Fig. 1 (a) Word-organized single crossed film cryotron cell array with current stretch sensing.

(b) Equivalent circuit during read-out.

(c) Current stretching circuit.

the superconducting path indicated by the broken line in Fig. 1(a). The equivalent circuit for this process is shown in Fig. 1(b).

This circuit is seen to be identical to the circuit of Fig. 1(c) if L_f is replaced by $L_a + L_d$ where L_a is the input inductance of the amplifier plus any interconnection inductance and L_d is the digit line inductance.

Thus the final amplifier current I_a equals:

$$I_a = \frac{L_c}{L_a + L_d + L_c} I_c \quad \dots(1)$$

where I_c is the initial current stored in the memory cell.

The reset c.f.c. of Fig. 1(a) serves two purposes: (1) to quench I_f when desired, and (2) to prevent the digit drive current from being shunted through the sense amplifier input when information is being written into a cell. The sensing process involves integration of the bit line voltage by a superconducting amplifier. One of the advantages of current stretch sensing is that any interference produced by inductive or capacitive coupling between word and bit line will be integrated out to zero. Another advantage is that, as shown by the derivation of eqn. (1), the rise time of the word line pulse is unimportant. That is, it can be made long compared to the cell time constant without affecting the amplitude of the read-out current.

To estimate the magnitude of the sense current in a practical case, assume a word-organized structure, as shown in Fig. 11, with 1000 cells per bit line. Let the cell inductance L_c equal 8 nH, let the bit line inductance be 2 nH per cell, and assume that the amplifier and interconnection inductance L_a is equal to the total bit line inductance. If the cell stored current is 100 mA, then from eqn. (1) the persistent amplifier input current is

$$I_a = \frac{8}{2000 + 2000 + 8} (100 \text{ mA}) = 0.2 \text{ mA.}$$

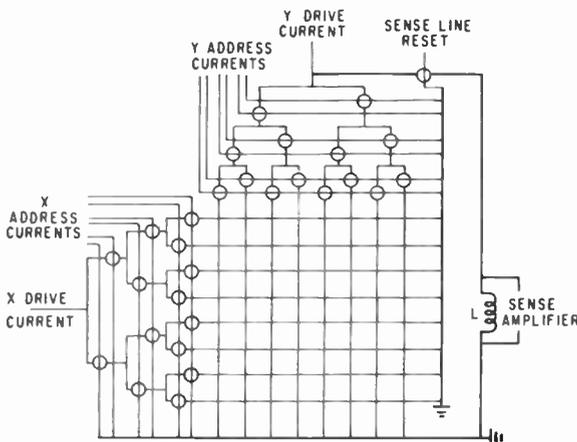


Fig. 2. Current stretch sensing for a bit-organized memory.

This can readily be detected by a c.f.c. amplifier³ of several megahertz bandwidth.

Current stretching can be used in a bit-organized memory configuration as shown in Fig. 2. Here an 8 by 8 array is shown with each cell represented as a crossover of X and Y lines and with the Y line being used for read-out. To read out a particular cell, current is first sent through the associated Y line, with the sense line reset cryotron held resistive to prevent shunt current passing through the sense amplifier. This cryotron is then made superconducting and current is sent through the selected X line. This produces a voltage across the selected cell, which in turn induces a persistent current through the superconducting path of the Y line cryotron tree and the sense amplifier. The capacitive loading of the Y line by the Y cryotron tree will not affect the current amplitude.

Equation (1) shows that use of the current stretching scheme requires that $(L_a + L_d)$ be kept as small as possible, in order to maximize the sensed current I_a . Hence all the storage cells described below are designed so as to keep the digit or Y line inductance L_d as small as possible compared to the cell inductance L_c .

3. Crossed Film Cryotron Memory Cells

3.1. A Three-cryotron Cell⁴ (Cell No. 1)

A three-c.f.c. cell with non-destructive read-out for use in a word-organized memory is shown in Fig. 3.† By simultaneously pulsing the X and Y write lines and terminating the Y write pulse last, current is stored in the loop of the Y write line. To read the cell, the X and Y read lines are simultaneously pulsed, creating a d.c. voltage on the Y read line if a current is present in the cell.

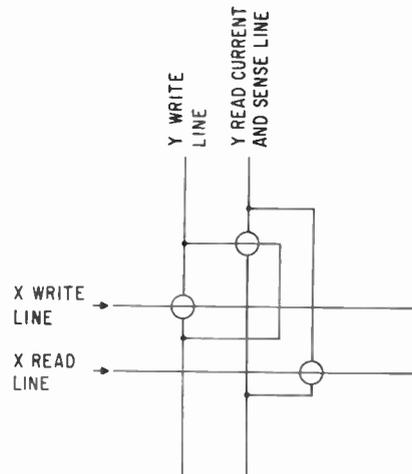


Fig. 3. Three-cryotron memory cell (cell No. 1).

† The presence of a superconducting ground plane is assumed, but not shown, in this and all other cells described in this paper.

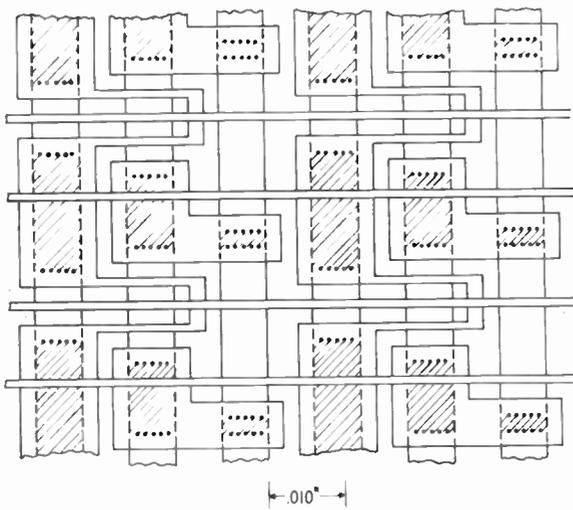


Fig. 4. Layout of three-cryotron cell array.

It has been generally accepted that this type of cell is neither sufficiently fast nor dense enough to be competitive in large memories. It has been included here, however, for completeness. Figure 4 shows a possible layout for such a cell, to be compared with other cells discussed later. The three vertical lines, equal in width, are tin (or indium) gates. All other lines are made of lead. The cross-hatched regions are Pb-Sn contact areas. With a 0.001 in-wide control and a 0.006 in-wide gate, the cell geometry of Fig. 4 occupies an area of 0.024 in \times 0.036 in, or 1150 cells/in².

To illustrate the dependence of cell area on fabrication precision, the area of this cell and of the cells described below are expressed in terms of the following parameters, which are illustrated in Fig. 5.

- w_g c.f.c. gate width
- w_c c.f.c. control width
- c minimum length of Pb-Sn contact area for a superconducting joint
- m minimum spacing between two non-overlapping films
- r minimum overlap of Pb overlying a gate
- r_i minimum overlap of Pb over a contact area-insulation interface.

Specific cell dimensions are given in Table 1. Other measures of cell density such as the number of insulated crossovers, the number of 'T' junctions, and the number of Pb-Sn contact areas are also given.

The time constant of the cell can be calculated from the well-known formulae for gate resistance and control inductance.

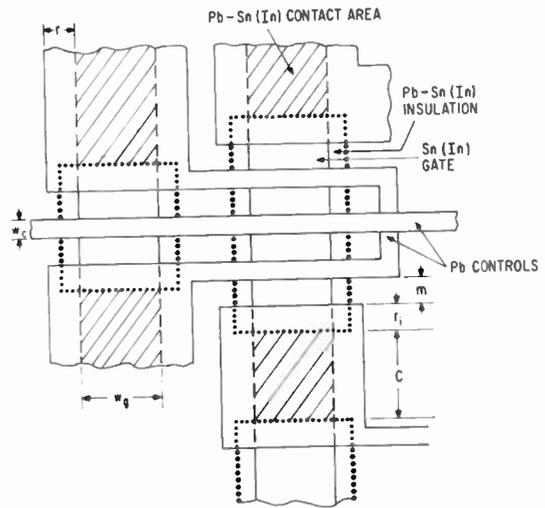


Fig. 5. Hypothetical layout to illustrate fabrication parameters (lead-lead insulation omitted).

Assuming the above cell size, the time constant of writing into the cell is 70 ns for a 5000 Å tin gate with a resistivity of 0.2 $\mu\Omega$ -cm and an insulation thickness of 10 000 Å. For reading, the time constant is 10 ns. The operative time constant for reading may be determined by other factors, such as the time constants of the selection circuits, the time constant of the sensing circuit, or the decay time of interference cross-talk from the drive-in line read pulses.

3.2. Double Cryotron Cell (Cell No. 2)

A double cryotron cell suitable for bit-organized operation is shown in Fig. 6, together with operating waveforms. The cell is read out destructively and the Y line is used for sensing as well as driving. If this cell is constructed from crossed film cryotrons, the Y line will contain a large number of controls and will have a correspondingly high inductance. This will greatly reduce the output signal if current stretching is used. The Y line inductance can be reduced by replacing the c.f.c. whose control is in the Y line with an in-line cryotron. Such an arrangement is shown in the layout of Fig. 7. As shown here, the cell is 0.021 in \times 0.031 in or about 1540 cells/in². In terms of our generalized parameters, the dimensions are $(w_g + 5w_c + 2r + 4m) \times (4m + b_{r_i} + 4c + w_c + L)$ where L is the length of the in-line cryotron gate, where the width of the narrow portion of the Y line memory loop is taken to be w_c and the width of films connecting to films of width w_c is arbitrarily taken to be $2w_c$.

Since the in-line cryotron always has a d.c. gain less than unity and since the c.f.c. gate must carry the control current of the in-line cryotron, it will be necessary to alter the structure of Fig. 7 slightly, by

making the in-line cryotron gate and control narrower than the c.f.c. gate. The exact geometry will depend on the gate thickness and the temperature of operation. In such a design, the c.f.c. will have sufficient gain for itself and the in-line cryotron. An alternative is to add a separate bias control to the in-line cryotron or apply a bias field to the entire plate. Besides the additional complexity that this would add, it might also interfere with the c.f.c. operation and hence is not desirable.

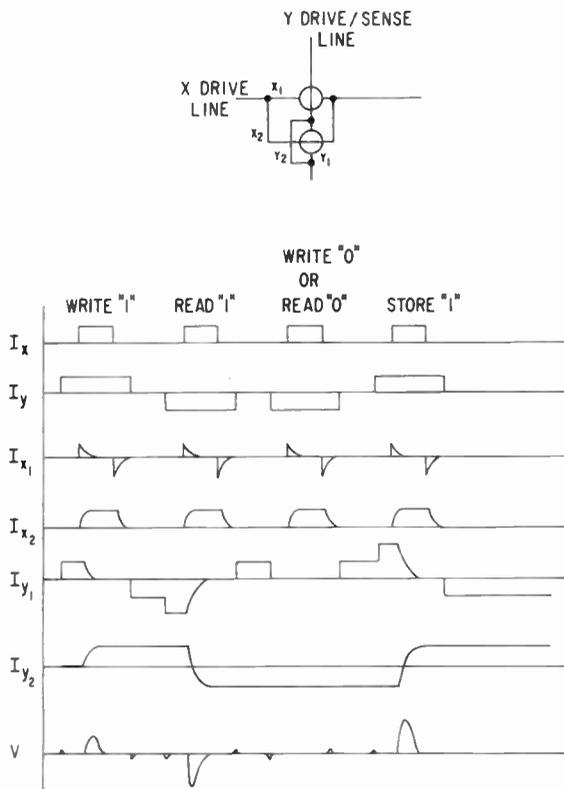


Fig. 6. A double cryotron memory cell and waveforms (cell No. 2).

With the geometry of Fig. 7, using a tin gate thickness† of $1\ \mu\text{m}$, a resistivity of $0.2\ \mu\ \Omega\text{-cm}$ and an effective shield insulation thickness of $1\ \mu\text{m}$, the time constant of the c.f.c. loop is about 60 ns and that of the in-line loop 20 ns. Supercooling effects associated with tin films thicker than the penetration depth have not been considered.

† The choice of gate thickness has been determined by the consideration that the in-line cryotron gate must be thicker than the penetration depth, and by the desirability of making the in-line and crossed film cryotron gates the same thickness since the geometry permits them to be deposited simultaneously.

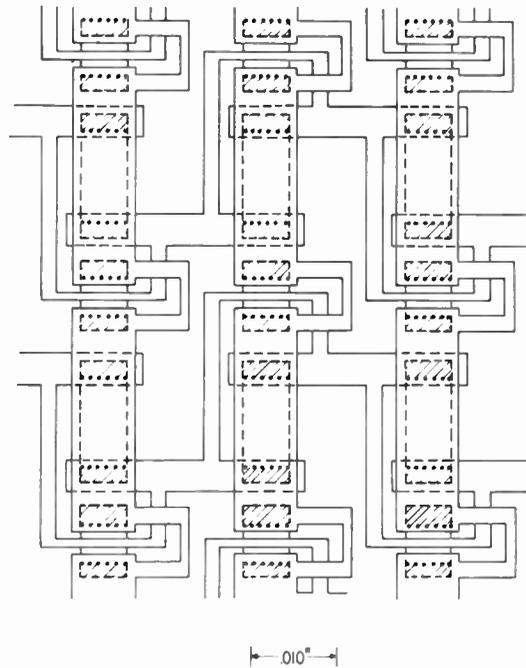


Fig. 7. Layout of double cryotron cell array.

3.3. A Low-inductance Double Cryotron Cell (Cell No. 3)

Another possibility of obtaining a low-inductance sense line using the basic cell of Fig. 6 is to remove the c.f.c. controls from the sense line and put them in a second vertical line. This is shown in the array of Fig. 8. The operating waveforms will also have to be

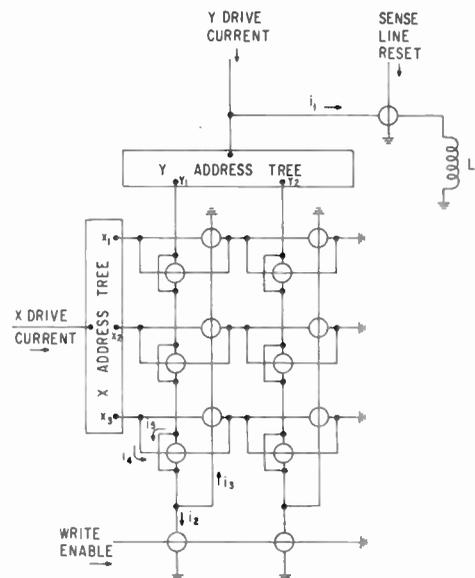


Fig. 8. A double cryotron cell in a bit-organized array (cell No. 3).

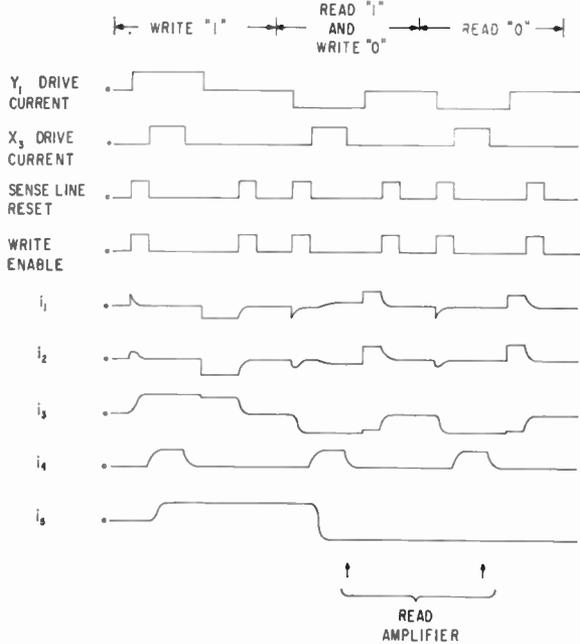


Fig. 9. Operating waveforms for cell No. 3.

modified. Figure 9 shows one of several possible modes of operating the cell. The general principle is to switch the Y drive current into the control line using the 'write enable' c.f.c. During sensing, the 'write enable' c.f.c. is superconducting, allowing the sense current to flow in the low-inductance path that

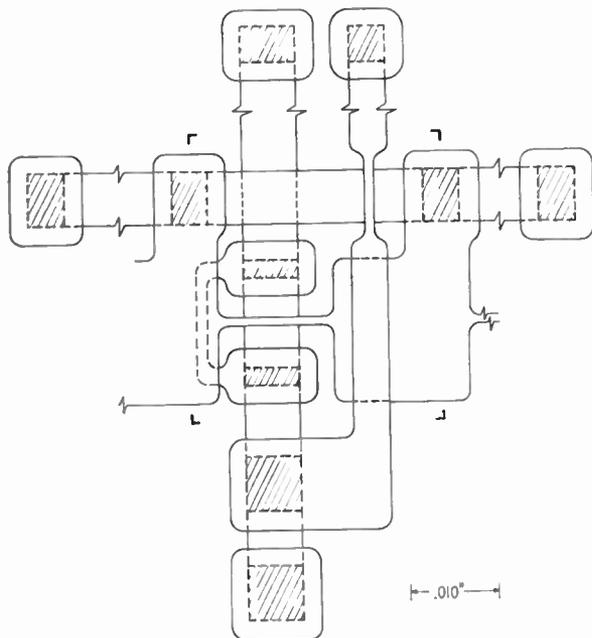


Fig. 10. Layout of cell No. 3.

bypasses the Y line c.f.c. controls. The 'sense line reset' c.f.c. keeps the Y drive current out of the amplifier input *L* and destroys the sense current after it is read out.

Figure 10 shows a possible layout for this cell designed by R. E. Joynson. It is 0.028 in × 0.030 in and has a density of about 1200 cells/in². A 256 by 256 array would conveniently fit on a 10-inch-square plate with associated address trees and an amplifier. In terms of fabrication parameters, the cell dimensions are $(w_g + 3r + 3m + c + \alpha w_c) \times (w_g + 4m + 4r_1 + w_c + 2c)$, where α is the ratio of the width of the non-active portions of the Y c.f.c. control line to w_c . This line is made wider between controls to lower the time constant of switching into the control line with the 'write enable' c.f.c. The cell has five insulated crossovers, four T junctions, and three Pb-gate contacts.

Assuming a resistivity of 0.2 μΩ-cm, insulation thickness 0.5 μm, and gate width 1.0 μm gives time constants of 18 ns for the switching loop and of 12 ns for the storage loop.

3.4. A Single Cryotron Cell (Cell No. 4)

A single cryotron cell that is very simple but requires a word-organized operation is shown in Fig. 11. The cell consists of a bottom layer of tin (or indium) arranged as a current storage loop and an overlying insulated-lead control line. Its operation has been explained in connection with Fig. 1. With a 0.002 in wide gate and a 0.001 in wide active control,

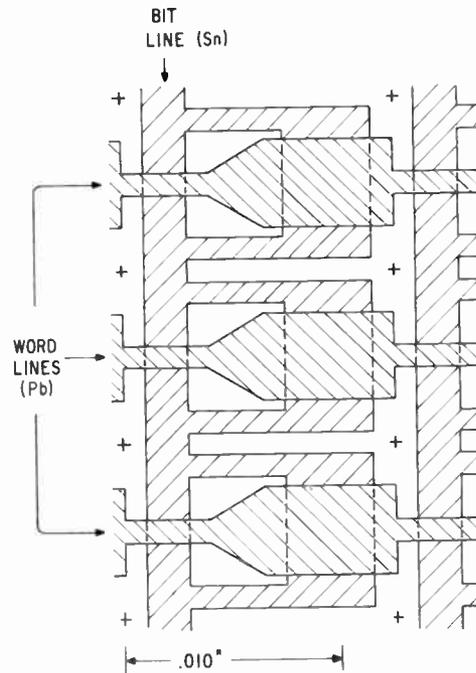


Fig. 11. Layout of cell No. 4.

the cell as shown would occupy 0.008 in \times 0.0125 in, giving 10 000 cells/in².

The simplicity of the cell is apparent since it only requires two insulated crossovers, two T-junctions, and *no* Pb-to-gate contacts. In terms of the fabrication parameters, the dimensions are $(w_g' + \beta w_g' + 2m + b) \times (2w_c + \gamma w_c + 2r - m)$, where β and γ are the multiplying factors for the gate and control necessary to create an inactive crossover in the right-hand leg of the cell, b is the length of the high-inductance portion of the memory loop, and w_g' is the gate width.

The maximum gate current in this cell is determined by the narrow portion of storage loop that is made equal to w_c . The gate in the Y line that is used for switching is made as wide as possible to reduce the inductance of the sensing circuit. It can be shown that this produces an overall increase in the sensing signal compared to a design where the gate has the same width as the narrow portion of the storage loop, or compared to a design where the storage loop leg width is increased to that of the gate.

An even denser arrangement⁵ of this circuit is shown in the four-cell array of Fig. 12, where the two halves of the storage loop are superimposed with the control sandwiched between them. With the geometry shown in Fig. 12, the cell area is only 0.005 in \times 0.007 in, giving 28 600 cells/in². With this cell, however, the two halves of the memory loop are made from different materials, and hence there is the problem of making superconducting Pb-Sn contacts. The time constant of this cell is proportional to the insulation thickness underneath the Pb leg of the storage cell. If this insulation is made thicker than a few microns, its edges would have to be 'sloped' so as to allow the Pb leg of the storage cell to 'climb' over it. These sloping edges will reduce the effective cell density. Ahrons⁵ used an insulation thickness of 12 μ m and obtained a 1 mV sense signal lasting for 200 ns. However, this required a cell 0.200 in long driven with 2 A word and digit currents! As in any word-organized memory, there exists the problem of plate-to-plate interconnections for all the digit lines.

4. The Interconnection Problem

Table 1 shows that the estimated cryotron cell density is highest for cells that must be used in a word-organized arrangement. This poses a severe interconnection problem, since although an address tree can be used for the word line, each *bit* line of a word-organized structure must be directly connected to the corresponding bit lines of the adjoining memory plates. For a 100-bit word, for instance, there must be at least 200-bit line interplate connections. For optimum memory performance, any direct connection between memory plates should be not only completely

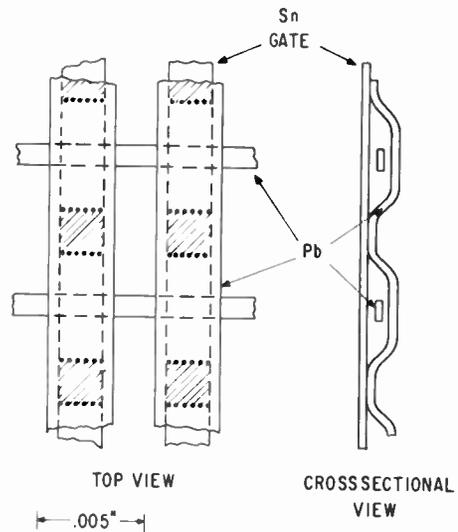


Fig. 12. High density arrangement of the one-cryotron cell (cell No. 5).

superconducting but also shielded. If the current stretching circuit is used, with one amplifier for each bit line in the memory, it is essential that both of these requirements be met.

Two ways of handling this problem are with (1) plug-in boards and (2) transformers. With the plug-in board technique, connecting pins would stick out of the edge of the memory plate wherever it is desired to make a connection. The films of the memory would be deposited over these pins in such a way as to make superconducting contacts where needed, and yet keep the circuit isolated from the shield plane, but with a low-inductance interconnection. The memory plates are then plugged into a master board in such a way as to make superconducting contacts. The shielded superconducting connectors between the individual memory plates are printed on the master plate of the memory. If current stretching is used, the bit lines of the various plates should be connected in parallel during sensing in order to reduce the inductance of the sense circuit. However, the bit lines must be serially connected during writing. These conflicting requirements will necessitate doubling the already large number of bit line interconnections and adding one or more switching c.f.c.'s to each bit line on every plate. (If a voltage sense could be used, the additional interconnections and the switching c.f.c.'s could be eliminated.)

A superconducting transformer effectively achieves the same result as superconducting contacts between the memory plate and the master board, since superconducting transformers have no lower cut-off frequency and have been demonstrated to transform direct currents. A transformer coupled memory

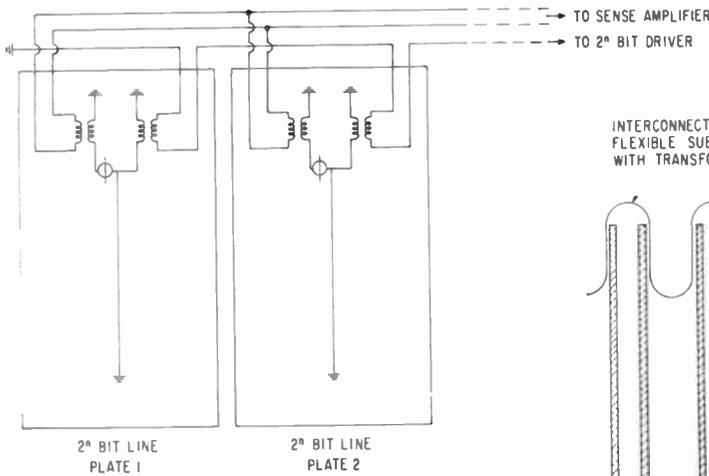


Fig. 13. Transformer coupled sensing system.

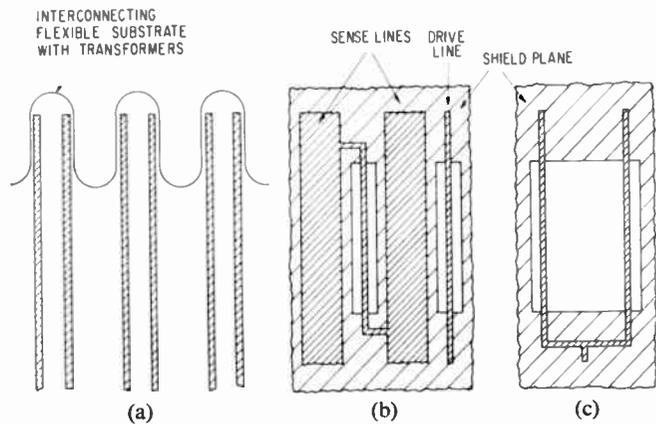


Fig. 14. Interconnection and transformer arrangement.

arrangement is shown in Fig. 13. Two transformers are used per digit line. One connects the corresponding bit lines on different plates in series for driving; the other connects them in parallel for sensing. One cryotron per digit line is required to perform this switching function.

When using a transformer coupled system, the memory plates might be interconnected by a flexible substrate carrying shielded lines, as shown in Fig. 14(a). The transformers themselves would consist of lines situated above holes in the shielding of the flexible interconnection (Fig. 14(b)). These are superimposed with matching lines above holes in the memory plate ground plane, shown in Fig. 14(c). These transformers need not be packed as densely as the bits on a word line. For instance, if 100-bit words are used on a 10-inch plate, each pair of transformers could occupy 0.1 inch of plate perimeter.

4.1. Transformer Design

Figure 15(a) shows a cross-section of a film transformer, and Figs. 15(b) and (c) show, respectively, the actual and equivalent circuits of a superconducting (i.e., resistance-less) transformer at frequencies where self-capacitance can be neglected. In Fig. 15(c), L_1 are the primary and secondary leakage inductances. These may be taken as equal to one another owing to the symmetry of the system. M is the mutual inductance.

Inspection of Fig. 15(c) shows that the load current produced by unit primary current is

$$T = \frac{M}{M + L_{load} + L_l} \dots\dots(2)$$

For T to be close to unity we must have

$$M \gg (L_{load} + L_l) \dots\dots(3)$$

The three quantities given in eqn. (3) can be calculated as follows. The load inductance L_{load} is that of a memory drive line of length l , width w , situated at a distance t above a ground plane. This is given by the well-known formula

$$L_{load} = 4\pi \frac{l}{w} t \times 10^{-9} \text{ H} \dots\dots(4)$$

(All dimensions in cm and $l \gg w \gg t$.) Figure 15(c) shows that the self-inductance of the transformer primary with the secondary short-circuited approximates $2L_1$ for $M \gg L_l$. But the self-inductance of the transformer primary with the secondary short-circuited is almost the same as that of a transmission line of the same dimensions above a complete, i.e. hole-free, ground plane. Hence we can use eqn. (4) to obtain

$$2L_1 = 4\pi \frac{l'}{w'} t' \times 10^{-9} \text{ H} \dots\dots(5)$$

where l' , w' , and t' are, respectively, the length, width, and separation from the secondary, of the transformer primary in cm. With good coupling between primary and secondary, the mutual inductance M is nearly equal to the self-inductance of the primary with the secondary open circuited, i.e., to the self-inductance of a line of width w in the centre of a ground plane hole of width $2g$. For this we may use the formula for the self-inductance of a round line of length l and diameter w , whose centre is a distance g above a ground plane, i.e.

$$M \simeq 2l \left[\ln \frac{4g}{w} + \frac{1}{4} \right] \times 10^{-9} \text{ H} \dots\dots(6)$$

We now must calculate the transformer dimensions which make $M \gg L_{load} + L_l$. Assuming a load con-

sisting of a 10-inch drive line, 0.002 inch wide, with a ground spacing of 0.5 μm, eqn. (4) gives

$$L_{load} = 3.14 \times 10^{-9} \text{ H.}$$

Assuming a transformer with dimensions

$$l' = 1 \text{ inch}$$

$$w' = 0.001 \text{ inch}$$

and

$$t' = 1.0 \text{ } \mu\text{m}$$

$$g = 0.00175 \text{ inch,}$$

eqns. (5) and (6) give

$$L_l \approx 10^{-9} \text{ H}$$

and

$$M \approx 10^{-8} \text{ H.}$$

These values of M and L_l satisfy eqn. (3). Hence the above dimensions provide a suitable transformer to efficiently drive a 10-inch memory digit line. It can be shown that the same size transformer will serve for sensing.

Reference to Fig. 13 shows that, during sensing, one sense transformer drives the sense amplifier shunted by all the rest of the sense transformer primaries. Assuming a word-organized memory consisting of 100 plates, this implies that the amplifier input is shunted by an inductance of $M/99$. To avoid an intolerable loss of signal, the sense amplifier input

must be kept comparable to or smaller than $M/100$. With the transformer arrangement of Fig. 14(a) and assuming a 100-plate memory, each sense transformer on the plate farthest away from the sense amplifier plate has to drive a 100-inch length of 'sense line'. (See Fig. 14(b).) To keep the inductance of this comparable to $M/100$, these sense lines will have to be made considerably wider than the 0.002-inch-wide drive lines, as shown in Fig. 14(b).

5. Discussion

On the basis of this study it appears that electronic limitations to the usable density of cryotron memory cells can be removed by the use of current stretch sensing. The attainable cell density is then limited only by the precision of fabrication technology. The 'proposed density' column of Table 1 shows the densities of the cells described, calculated in terms of parameter values that represent what are considered to be the limits of the present state-of-the-art. For the cells under discussion, these limits are determined by the minimum overlap distance between different metal films and thus between the photomasks used to form them. Table 1 shows that for *bit-organized* structures the anticipated maximum cell density lies between 1000 and 3000 cells/in².

A present-day liquid helium refrigerator with a working volume of 1 ft³ could hold a memory consisting of 100 plates, each with an area of 100 in². Assuming an average density of 2000 cells/in² leads to a memory with 2×10^7 bits. Since the refrigerator would probably cost between \$10 000 and \$50 000, the refrigeration cost would be between 0.05 and 0.25 cent/bit. However, 10^7 bit memories have already been built by using magnetic cores and are projected using magnetic films. It is not clear that any significant cost savings would be achieved by using a bit-organized cryotron memory, even though this has the advantage of being an integrated memory with the cryotron address trees fabricated on the same substrate as the cryotron memory cells.

Table 1 shows that estimated *word-organized* cryotron cell densities are a factor of 10 larger than bit-organized cell densities. Hence a 1 ft³ refrigerator could hold 10^8 bits, and the refrigerator cost would fall to well below 0.1 cent/bit. Such a cryogenic memory with its integrated memory address tree undoubtedly becomes competitive with magnetic memories of any type, especially since magnetic memories have drive-line and sense-line attenuation problems at this size.

However, it has been shown that word-organized cryotron memories have their own engineering problem that is not encountered by bit-organized structures: that of providing each memory plate with several hundred interconnections. Furthermore, if

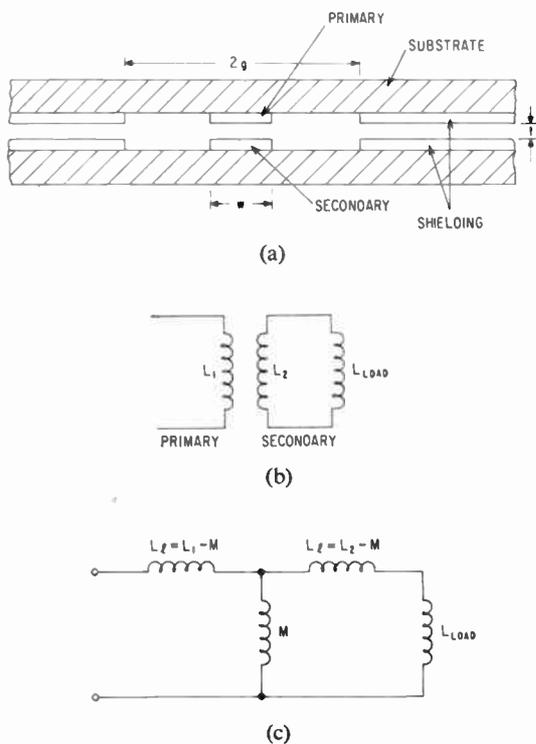


Fig. 15. Transformer cross-section and equivalent circuit.

current stretch sensing is required, these interconnections must be superconducting, must have low inductance, and must be arranged so that the bit lines from the different plates can be driven in series and sensed in parallel. The feasibility of such an interconnection scheme remains to be demonstrated.

It is pertinent to compare the prospects of word-organized cryotron and bit-organized continuous film cryogenic memories. Continuous film memories, which are presently under consideration, employ a cell structure whose complexity is roughly comparable to that of the simpler structures considered above. In other words, the c.f.m. cell requires no contact between different metals but involves precise alignment of different metal layers. The proposed density for an initial c.f.m. is around 10 000 cells/in², which is no better than what should be attainable with word-organized cryotron cells. However, the c.f.m. has the advantage of bit organization and thus faces no interconnection problems. On the other hand, it has the disadvantage with respect to cryotron cells of having narrower tolerances. In conclusion, it would

appear that no easy choice can be made at this time between continuous films and cryotron memories. Both present problems which have not as yet been experimentally investigated.

6. References

1. H. H. Edwards and V. L. Newhouse, 'Analysis of the cryogenic continuous film memory', *Trans. Inst. Elect. Electronics Engrs on Magnetics*, MAG-1, No. 4, pp. 369-78, December 1965.
2. R. L. Garwin, *IBM J. Res. Develop.*, 1, p. 304, 1957. R. K. Richards, Symp. on Superconducting Computer Devices, London, 7th December 1964; also A. E. Slade, *Proc. Intern. Symp. on Techniques of Memories*, Paris, 5th-10th April 1965.
3. V. L. Newhouse and H. H. Edwards, 'An ultrasensitive linear cryotron amplifier', *Proc. I.E.E.E.*, 52, No. 10, pp. 1191-1206, October 1964.
4. M. K. Haynes, 'Cryotron storage, arithmetic and logical circuits', *Solid State Electronics*, 1, No. 4, p. 399, April 1960.
5. R. W. Ahrons, *Proc. Intermag. Conf.*, 1965. (Paper 9.1.)

Manuscript received by the Institution on 6th April 1966. (Paper No. 1102/C91.)

© The Institution of Electronic and Radio Engineers, 1967

A Computer Controlled Precision Film Scanner

By

C. A. BORDNER, Jr., Ph.D.,†

Professor

A. E. BRENNER, Ph.D.‡

AND

P. de BRUYNE,

B.Sc., C.Eng., A.M.I.E.R.E.‡

Summary: A flying-spot film scanner was developed, capable of a short sweep in any direction from any of the 1024×1024 precise positions on the face of a 5-inch c.r.t. The spot is imaged on a film with 1 : 1 optics.

Fine black lines on the film can be located with an accuracy of about $3 \mu\text{m}$. The width and degree of blackness of these are also recorded and these data are transferred to a computer on its demand.

The computer processes the data and provides fresh instructions for the scanner so that the following sweep vector can be optimally specified.

1. Introduction

Spark and bubble chambers are devices used to determine the paths of nuclear particles in high-energy physics experiments. Photographs of these paths or tracks as they are called, are made in large numbers and require careful analysis. Typical measurements require a precision down to $\pm 3 \mu\text{m}$ on 70 mm wide film. Two or more views are used so that three-dimensional information may be reconstructed. In some cases all views are imaged on one film frame. Each track may be precisely related in space to identification (fiducial) marks, hence the calibration is inherently simple.

Manual measurements tend to become unsuitable as the number of frames to be analysed becomes large. It has been estimated¹ that the total output from spark and bubble chambers located at high-energy particle accelerator sites will reach well over twelve million stereo pictures for 1966 and so automatic scanning and data reduction must be used to an increasing extent.

The system to be described aims to process of the order of one million pictures a year without the large capital expenditures commonly required for an automatic system of such accuracy. The device is essentially a computer-controlled flying-spot scanner which has random access to any point on the frame and can produce a sweep vector of variable length and direction from that point. It differs in this respect from a system described earlier² where the spot size was larger and no sweep vector was used. The accuracy is now of the order of 1 part in 3×10^4 whilst previously it was only 1 in 10^3 .

The system now combines the feature of high accuracy with that of flexibility under computer control. It makes use of some of the ideas used in several other systems.^{3,4,5}

A relatively small computer is used (a PDP-1 of Digital Equipment Corporation, Maynard, Mass.)

† Formerly at Harvard University Cyclotron Laboratory; now at Department of Physics, Colorado College, Colorado Springs, Colo.

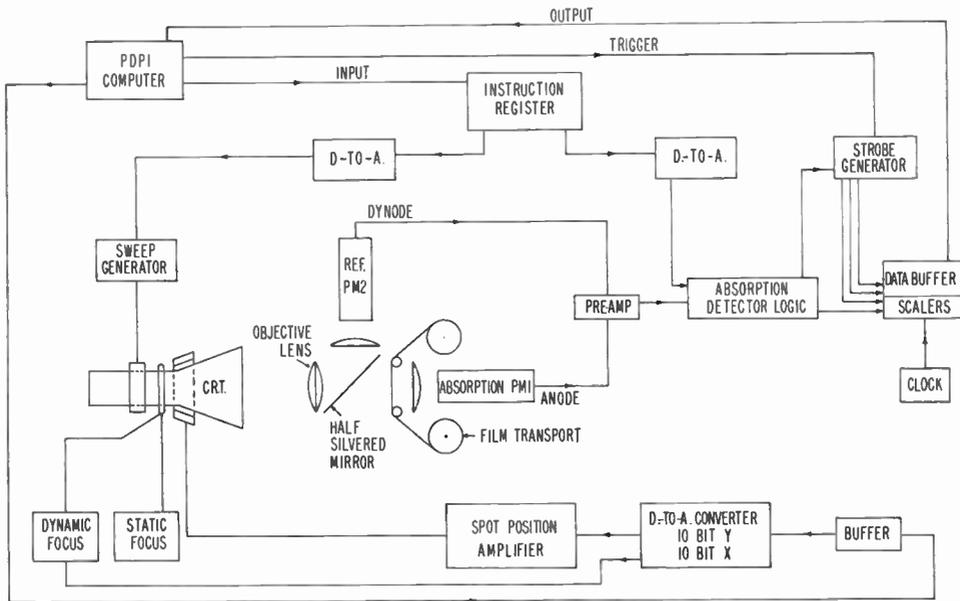
‡ Harvard University, Cyclotron Laboratory, Cambridge, Massachusetts.

together with a controller. This contains the interface electronic equipment to the input-output register of the computer, together with a scanning table which holds the electro-optical parts. The major components are identified in Fig. 1.

The computer controls starting position, inclination and length of the sweep of a $25 \mu\text{m}$ spot over the 10×10 cm active area of the c.r.t. This spot is imaged with 1 : 1 optics on a film. This film has data on it in the form of fine black lines which correspond to the tracks of nuclear particles in spark chambers. The useful duration of the sweep is $12.8 \mu\text{s}$. Thus by noting the time at which the spot crosses the film data, the precise position may be obtained. The sweep length selected corresponds to 127 clock counts. The absorption centre is thus calculated by the computer from the starting position, total length of the sweep and the location number in the buffer.

2. Electro-optical System

The flying-spot scanner was constructed using a Dumont KC2347 cathode-ray tube and a set of precision focus and deflection coils made by Celco of Mahwah, N.J. The objective lens is an Elgeet of $f1.9$, 15 cm, designed for 1 : 1 magnification and corrected for P16 phosphor. The spot-size on the c.r.t. was generally less than $30 \mu\text{m}$. Pin cushion and other distortions affecting the spot position are corrected by using many reference marks on the film, each serving its own local area. All detected positions are thus related to some of these marks by computer action. A half-silvered mirror was made of $25 \mu\text{m}$ mylar and placed at a 45° angle behind the lens (Fig. 1). A brightness reference signal is obtained from a photomultiplier viewing the reflected light and the output is subtracted from the absorption signal of a photomultiplier placed behind the film. The photomultiplier voltages are adjusted so that their combined output is about zero at positions of absorption (black film). The improvement of this arrangement over the use of only one photomultiplier is considerable and is essential to a system analysing spark chamber pictures using the film negative.



The computer supplies a digital address in x and y which controls the spot position via a converter and amplifier. It also supplies the instruction register with a display instruction which specifies the sweep length and direction to be produced by the sweep generator. A trigger pulse is then given which initiates the sweep. The detected signal is digitized in the absorption detector logic and scalers and is presented to the computer via the data buffer.

Fig. 1. Principal block diagram.

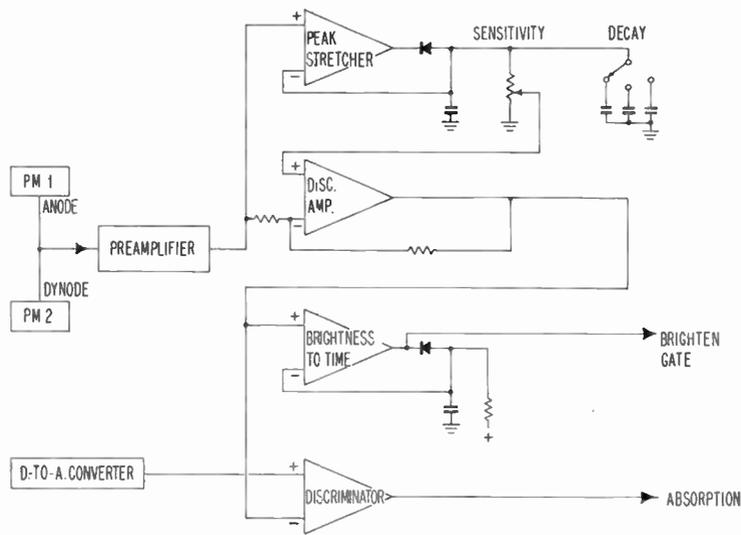


Fig. 2. Absorption detector logic.

3. Discriminator

A single photomultiplier as is used in mechanical flying-spot scanners cannot be used with a c.r.t. scanner because of brightness variations of over 3 to 1 at various positions on the screen. The system chosen not only looks at the film differentially with two photomultipliers but also uses a variable absorption threshold so as to overcome the following problem.

When moving to various positions on the film frame, the fraction of light reaching one photomultiplier changes with respect to the fraction reaching the other one. This is due not only to data, but also to changes in background film density and to imperfections in the light gathering system. A simple difference signal between the two photomultiplier outputs is thus not always a good measure of film data. With spark chamber pictures, it is better to refer this difference to one in the immediate vicinity, showing a minimum of absorption. In practice this can be accomplished by a peak stretcher which stores a voltage corresponding to the peak transmission of previously scanned points (see Fig. 2).

A threshold is then set at some fraction of this voltage and the instantaneous value of the absorption signal is continuously compared with this level in a linear discriminating amplifier. The output of this amplifier is compared with a 'sensitivity' level set up by the computer in three bits of the instruction register. There are, hence, eight possible sensitivity levels of the discriminator.

The output of this unit is a logic level called the absorption level. The above system allows a uniformly high sensitivity over the whole film in the presence of large c.r.t. brightness variations and unequal lens vignetting of the two photomultiplier channels.

4. Width, Centre Position of Absorption Level and Brightness Information

The duration of the absorption level is used in gating a width scaler which accumulates a number of 10 MHz counts corresponding to the width of the mark on the film. During the sweep a 10 MHz clock is counting into a 7-bit scaler. When the discriminator output starts to show a film absorption, the contents of the scaler are transferred into a half-speed scaler counting at 5 MHz. The degree of absorption is recorded for each absorption peak which produces a discriminator output. An analogue to digital time gate generator is connected to the output of the discriminating amplifier and provides an output of 0 to 0.8 μ s duration depending on the depth of absorption (brightness of spark). The brightness counter is allowed to count at 10 MHz for this time allowing 8 possible brightness levels to be recorded.

At the end of the absorption, the counting of the width and the half-speed position counter stop. The contents of all these counters and several other data flip-flops are allowed to transfer into a buffer 2 and subsequently into a final buffer 3 provided they were not already in use. A total of 3 separate absorptions may thus be recorded in a single sweep period.

5. Operations Controlled by the Computer

(a) Major spot position: The computer produces a digital address in x and y coordinates which, via an accurate converter and amplifier, positions the blanked beam of the c.r.t. on any one of 1024×1024 precise starting positions on the c.r.t. face. The starting position may thus be moved in steps of $\pm n \times 75 \mu$ m where n is an integer. The active area on the c.r.t. is about 77×77 mm.

(b) Before unblanking the beam, an 18-bit buffer called an instruction register is set by the computer:

bit 0, 1 and 2 control the discriminator sensitivity,
bits 3–8 specify the x sweep amplitude 0 to 2000 μ m.

The active sweep is always of 12.8 μ s duration. If bits 3–8 are all asserting a 2000 μ m sweep is obtained and the spot speed is maximum. If none are asserting, the spot will stay stationary. Final precision measurements are usually made with a 250 or 125 μ m long sweep. Elsewhere in this paper it is seen that the detector logic allows 1, 2 or 3 locations out of 127 to be recorded along the sweep whatever its length may be.

bit 9 controls the x sweep direction,

bits 10–15 specify the y sweep amplitude 0 to 2000 μ m,

bit 16 controls the y sweep direction,

bit 17 enables the high voltage for the photomultiplier and enables a subsequent beam unblanking.

(c) When the above instructions have been given, the computer may give a trigger pulse which initiates a 16 μ s sweep (see Figs. 3 and 4). The length, direction and starting point have previously been specified. At the end of the sweep the output buffer will be interrogated and any data that have been stored will be transferred to the computer. In order to sample the peak brightness before the sweep starts, the spot is actually backed off 10% of the sweep length and allowed to sweep through the position of zero deflection and up to its maximum value in one continuous sweep. The beam is unblanked before the position of zero deflection is passed and the discriminator will thus retrieve a reference peak brightness before the scaler starts. (See Figs. 5 and 6).

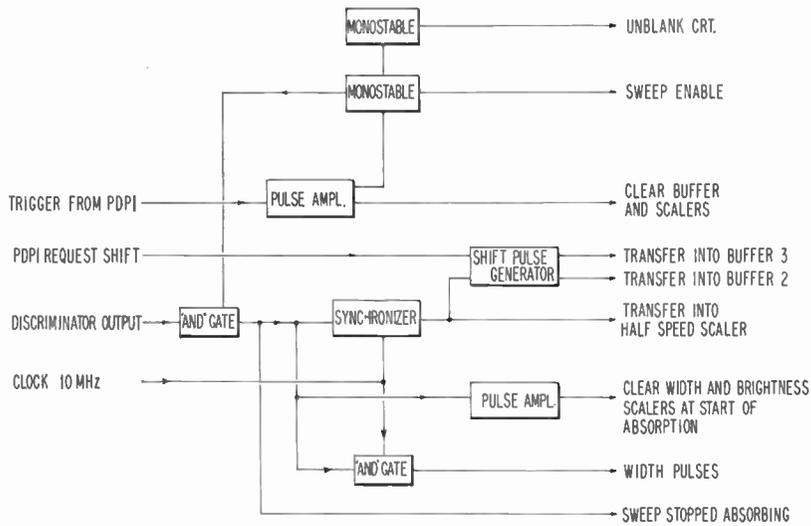


Fig. 3. Strobe generator.

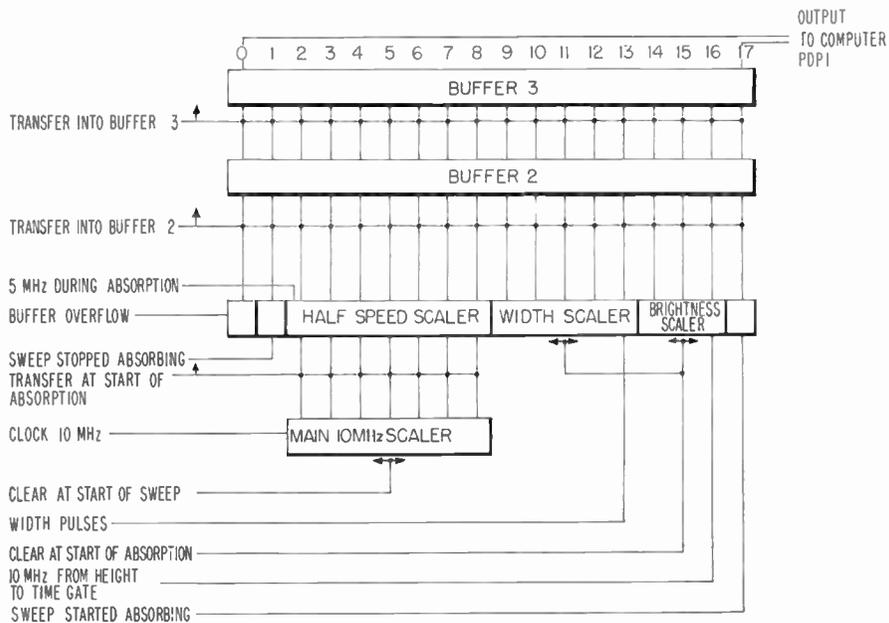


Fig. 4. Data buffer and scalers.

6. Interface Output to the Computer

The buffer 3 presents its output via cable drivers to the computer. The 18 bits are designated:

- bit 0 (a) There is a data word following in the buffer 2, or...
- (b) In the case that two words have already transferred, it signifies that there were one or more words which overflowed the buffer.

- bit 1 The sweep stopped when the discriminator was showing an absorption level.
- bit 2-8 The centre position of the absorption level referred to the start of the sweep in clock counts (7 bits, any one of 127 locations).
- bit 9 width overflow,
- bit 10-13 width in basic clock counts,
- bit 10-16 brightness or degree of absorption,
- bit 17 not used.

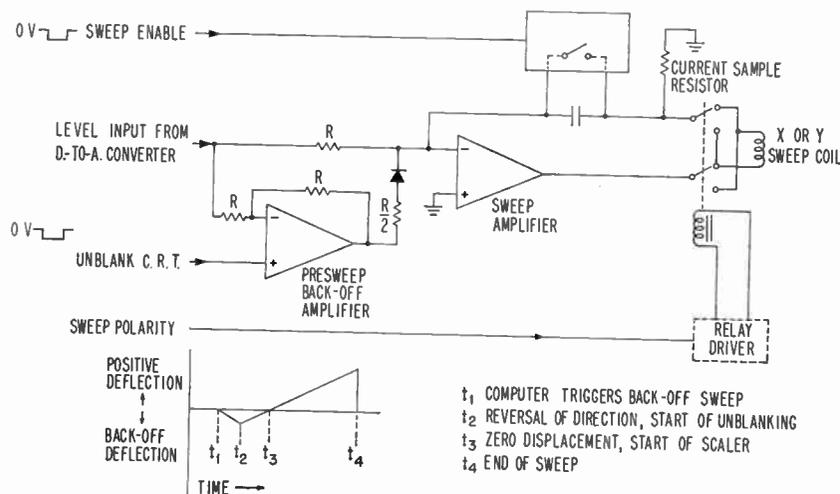


Fig. 5. Sweep generator and back-off amplifier.

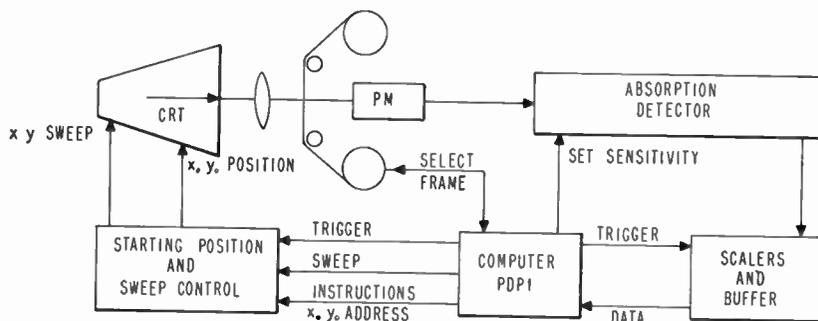
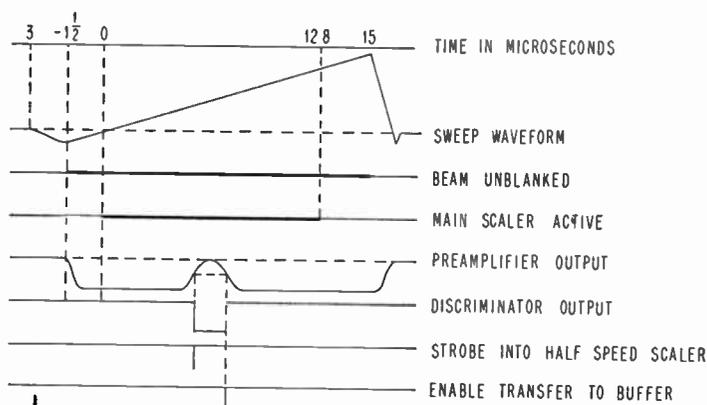


Fig. 6. Typical computer-controlled timing sequence:

- (1) Select frame and wait for ready reply.
- (2) Address starting position as one point on 1024×1024 grid.
- (3) Set instruction register for sweep length, direction and sensitivity.
- (4) Trigger sweep-circuits resulting in sweep waveforms shown.
- (5) Read-out buffer 3.
- (6) Calculate next function and execute it.

When the first word has been transferred, the computer will give a return pulse if bit 0 was asserting. This pulse will clear the buffer 3 allowing the information in buffer 2 to take its place. This will be repeated once more for buffer 1 allowing all three words to be transferred. The computer will then calculate a new instruction and the whole cycle will repeat itself.

If no film data are found or if too much noise is present, the instructions may call for a change in beam brightness, sweep length or discriminator sensitivity. If the sweep begins or ends in an absorbing area, the starting position may be changed by any number of steps of 75 μm , or a new sweep length selected using a new instruction word.

A useful test program which has been written, displays an enlarged version of any area of the film on a monitor display c.r.t. The effect of changes in sensitivity, optical and electrical focus controls may be readily observed on a monitor oscilloscope. Scanning in this case is carried out as a large number of partially overlapping horizontal and vertical lines. The centre and width of the detected test pattern lines are stored in computer memory. This information is re-displayed at any required magnification.

7. Computer Programs

When writing instructions for the computer, many facts have to be borne in mind so as to achieve both speed and accuracy in the measurements. The film which, in this application, consists of high-energy physics spark chamber photographs has 4 major identifying location marks. These are scanned, each with a horizontal and vertical line pattern of 2 mm to locate their positions. The secondary marks are located next in positions referred to the 4 major marks. There may be a dozen or more of these marks which are located with an accuracy of 2 to 3 μm using 250 or 125 μm sweep lengths. Before making any final accuracy measurements, the blanked spot must remain near the area of measurement for several milliseconds to allow the amplifiers to settle and allow induced currents near the deflection coils to decay.

Actual hysteresis of the spot position due to magnetic remanence was found to be negligible but periodic checks may be made and dual measurement carried out when necessary. As the film data are always near one of the minor location marks, the appropriate measurements can easily be made so that hysteresis will not affect the accuracy.

8. Results

Tests have indicated that the stability of spot position is better than 3 μm over a period of minutes. All film information is always referred to identification marks close to the film data on the film. These are measured less than one second before so there are no great demands on the long-term stability. The minimum absorption widths are about 30 μm because of the spot diameter, but even widths of 100 μm can be located to an accuracy of $\pm 3 \mu\text{m}$. As the circuitry always finds the centre of the absorption, the recorded position is independent of the sensitivity used, even though the recorded width may vary from 20 to more than 100 μm .

The instrument has been found very flexible in adapting to various types of photographs and can undoubtedly be used whenever large numbers of scientific pictures have to be analysed with a high degree of precision.

9. Acknowledgments

The authors wish to express their thanks to Mr. B. F. Wadsworth of Massachusetts Institute of Technology, for valuable discussions and advice, and to Dr. C. Zajde for helpful comments regarding this paper.

10. References

1. B. F. Wadsworth, 'PEPR—A developmental system for rapid processing of bubble chamber film', Conference on Filmed Data and Computers, June 1966, Boston, Mass., *Proc. Soc. Photogr. Instrum. Engrs*, § IX, pp. 1-14.
 2. P. de Bruyne, 'A light absorption detector for a computer controlled film scanner', *The Radio and Electronic Engineer*, 29, No. 5, p. 325, May 1965.
 3. I. A. Pless *et al.*, 'A precision encoding and pattern recognition system (PEPR)', Proceedings, XIth International Conference on High Energy Physics, Dubna, U.S.S.R., 1964.
 4. P. V. C. Hough and B. W. Powell, 'A method for fast analysis of bubble chamber pictures', *Nuovo Cimento*, 18, pp. 1184-91, August 1960.
 5. H. Rudloe, M. Deutsch and T. Merrill, 'PIP. A photo interpretive program for the analysis of spark chamber data', *Commun. Assoc. Computing Machinery*, 6, No. 6, p. 332, June 1963.
- M. Deutsch, 'A spark chamber automatic scanning system', Proceedings of the Conference on Photon-interactions in the BeV Energy Range, Massachusetts Institute of Technology and Laboratory of Nuclear Science, 1963. (Unpublished.)

Manuscript first received by the Institution on 4th July 1966 and in final form on 16th December 1966. (Paper No. 1103/C92).

© The Institution of Electronic and Radio Engineers, 1967

Analogue Computer Study of the Effect of Input Spectrum on a Hill-climbing Optimizer

By

I. COCHRANE,
B.Sc., Ph.D., C.Eng., M.I.E.E.,†

A. PERETTI,
B.Sc., A.R.C.S.T.†

AND

A. B. GARDINER,
B.Sc., A.R.C.S.T.†

Summary: One of the techniques used in the optimization of controlled plant is that of hill-climbing. This technique is most usefully employed on-line and, therefore, must operate with the normal random operating signals. The performance (as opposed to economic) criterion of the plant's operation is usually chosen to be the mean square error. This paper investigates the variations in the shape of the curves of mean square error against the forward, and rate feedback, gains of a generalized system as the bandwidth and sharpness of cut-off of the random signals are varied. Considering the signals as filtered white noise, it is shown that optimization becomes difficult if the filter pole is of order greater than three, and the filter pole must be positioned carefully relative to the system's open-loop poles, if optimization is to be achieved at all using both parameters.

List of Symbols

a	gain of rate feedback loop
D	normalized value of mean square error at a minimum
E	error transfer function
f	stabilizing-loop transfer function
F	forward transfer function
G	open-loop transfer function
H	closed-loop transfer function
K	gain of forward transfer function
K_c	gain of forward transfer function producing instability
m	order of denominator of F
N	value of input filter, or noise, pole
p	system pole values in F
P	normalized values of p
s	complex variable
t	time
x	order of noise pole
z	system zero in F
Z	normalized value of z
ε	system error, $(\theta_i - \theta_o)$
θ_d	disturbance signal
θ_i	input quantity
θ_o	output quantity
ω	angular frequency
$\bar{\varepsilon}^2$	mean square error

1. Introduction

In the design of self-optimizing control systems the method used broadly depends on whether the mathematical model of the system is known or not. If it is known sufficiently accurately, modal analysis may be used, but if little is known about the system dynamics, then an automatic on-line hill-climbing method¹ is most useful. As the hill-climbing process is carried out on-line (since a suitable model cannot be obtained), the signals going through the system are the normal plant operating signals, which are usually random in nature. These operating signals may thus be thought of as bandwidth-limited white noise, the bandwidth being determined by their source, or by their source and the plant time-constants.

In the analysis of hill-climbing optimization, it is common in the literature^{2, 3} to assume that there is either a quadratic or parabolic relationship between the chosen criterion of performance and each of the parameters of the system used for optimization. Due to the random nature of the operating signals, the criterion is normally the mean square of the error between desired and actual values of the controlled variable,^{2, 4} as shown in Fig. 1. However, the error is a function of both the operating or disturbance signals, and the plant dynamics. Therefore, it is pertinent to enquire under what conditions the mean square error (m.s.e.) does, in fact, go through a minimum value, as the effective plant dynamics are varied, and the validity of the quadratic assumptions made about m.s.e./parameter relationship.

Unfortunately, analytical evaluation of the m.s.e.^{4, 5} is of no help in predicting the shape of the m.s.e. 'hill', or whether a minimum value exists, for any but the simplest of systems and disturbance signals. Therefore, the problem was investigated experimen-

† Department of Electrical Engineering, University of Strathclyde, Glasgow.

tally, using analogue simulation of systems whose order varied from second to fifth, and white noise disturbance signals whose bandwidth and rate of cut-off were independently variable.

2. The Experimental System

The main units of the experimental system are shown in Fig. 1. The closed-loop system within the dotted lines comprises a linear forward transfer-function, $F(s)$, and a rate-feedback stabilizing loop $f(s)$. The gains of these two transfer-functions are used as the optimizing parameters. In computing the results, specific transfer-functions are assumed for $F(s)$, and the filter cut-off frequency referred to them.

Specifically, it is assumed that

$$F(s) = \frac{K_1(s+z)}{s(s+p_2)\dots(s+p_m)} \dots\dots(1)$$

$(2 \leq m \leq 5, m \text{ integer})$

To allow generalization, $F(s)$ is normalized by reducing the principal lag, say p_2 , to unity. Writing $s_1 = s/p_2$,

$$F(s_1) = \frac{K(s_1+Z)}{s_1(s_1+1)\dots(s_1+P_m)} \dots\dots(2)$$

where $P_r = p_r/p_2$, $Z = z/p_2$, and $K = K_1/(p_2)^{(m-1)}$. Similarly, for rate feedback, $f(s) = (a_1)s$ and $f(s_1) = (a_1/p_2)s_1 = (a)s_1$. Normalization is, in effect, a change of the time scale. The error transfer function is:

$$\frac{\varepsilon}{\theta_i}(s_1) = \frac{1}{1+G(s_1)} \dots\dots(3)$$

where $G(s_1)$ is the open-loop transfer function. Since eqn. (3) has the same denominator as the closed-loop transfer function, θ_o/θ_i , the root-locus for the error transfer function is identical with that of the closed-loop. Further, since the general shape of the root-locus depends only upon the number of open-loop poles, and not upon their relative spacings, the choice of the positions of the poles in $F(s_1)$ is arbitrary (with the exception of the poles at the origin and at

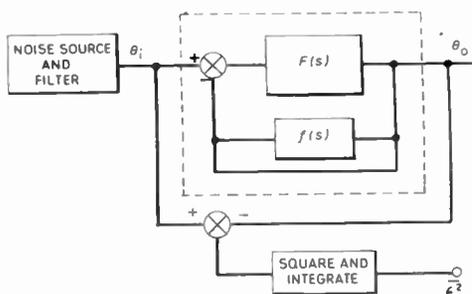


Fig. 1. Block diagram of the experimental simulation. The dotted line encloses the linear system to be optimized.

$s_1 = -1$). Thus, octave spacing is used, i.e. poles at $s_1 = 0, -1, -2, -4, -8$, the more negative poles being deleted as the order of the system is reduced. The effect of the zero, Z , in $F(s_1)$, is measured as the zero is moved from the outside of the pole pattern in towards the origin.

When rate feedback is included, its effect is to inject a zero into the root-locus, i.e. expanding eqn. (3),

$$\frac{\varepsilon}{\theta_i}(s_1) = \frac{1+s_1(a)F(s_1)}{1+a(s_1+1/a)F(s_1)} \dots\dots(4)$$

The result on the system of increasing a is not necessarily stabilizing,⁶ except for second- and third-order systems where it always is. For fourth- and higher-order systems, provided K is not so large that rate feedback alone can never produce stability, as the feedback gain a is increased the system will only be stable for a range of values of a . This is best shown by the inverse frequency plane plot of the situation (see Fig. 2).

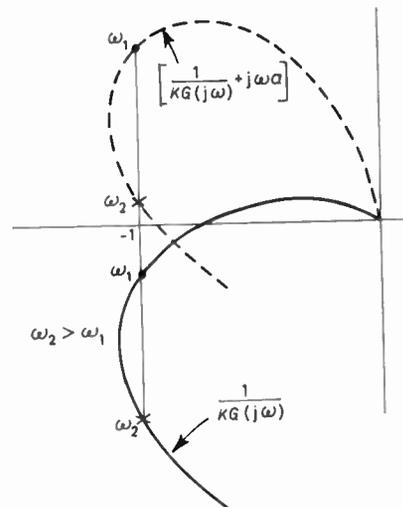


Fig. 2. Diagram showing that too little, or too much, rate feedback can produce an unstable system.

The mean square error is measured as a function of the two parameters, K and a .

In defining the spectrum of the disturbance signal, consideration must be given to practical situations. If the disturbance emanates from a source external to the system, its spectrum is independent of the system and eqn. (3) applies. Alternatively, if the signal appears at some point within the system, as in a regulator-type system operating with a fixed set-point, then if $G_1(s_1)$ is the portion of $G(s_1)$ between the entry of the signal and the system output, the spectrum is given by

$$\theta_d(s_1)G_1(s_1)/(1+G(s_1)) \dots\dots(5)$$

in which, if $\theta_d(s_1)$ is the actual disturbance, $\theta_d(s_1)G(s_1)$ may be looked on as an equivalent input. The latter has a frequency spectrum which is partly a function of the system itself. Therefore, a reasonable spectrum for the experimental input will be:

$$\theta_i(s_1) = \frac{1}{(s_1 + N)^x} \quad x \text{ integer} \quad \dots\dots(6)$$

where $N = n/p_2$ is the normalized filter pole. The value of x will be at least unity for the case of eqn. (3), and two or more for eqn. (5) allowing for lags in $G_1(s_1)$. In this investigation x lay in the range, $1 \leq x \leq 4$.

3. Simulation

The analogue simulation of the transfer function of the linear closed-loop system is achieved by standard means. However, the integration of the error squared produces a practical difficulty. If the disturbance signal is filtered white noise, the error after squaring, must be integrated for a very long period,^{7, 8} identical in each computing run. Also, due to the random amplitude distribution of the noise, its mean square value must be kept low to prevent peak overloading in the computer. Therefore, the small drift levels introduced by any form of analogue multiplier can have a material effect on the value of the integral.

Since the system is linear, these difficulties are overcome by using as an input signal the time function which is the transform of the required frequency spectrum. For the disturbance spectrum considered in eqn. (6):

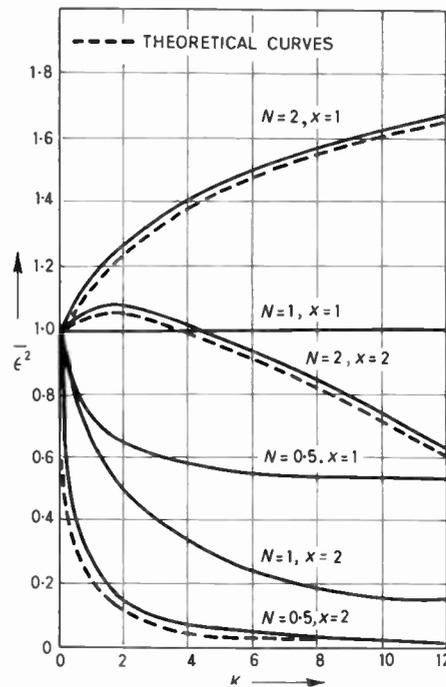
$$f(t_1) = \frac{1}{(x-1)!} \frac{d^{(x-1)}}{ds_1^{(x-1)}} e^{s_1 t_1} \Big|_{s_1 = -N} \\ = \frac{t_1^{(x-1)}}{(x-1)!} e^{-N t_1} \quad \dots\dots(7)$$

Since $f(t_1) = 0$ essentially for $N t_1 \simeq 10$, the integral of error squared will reach its final value in the time $t_1 \simeq 10/N$. This time can be made as short as convenient by time-scaling the entire simulation, and the amplitude of the disturbance can also be increased without fear of overloading. The analogue may now be used with logic elements as a hybrid computer for fully automatic solution.

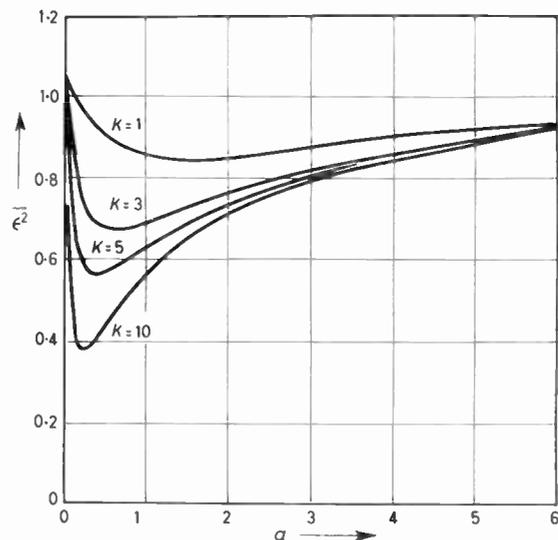
4. Experimental Results

To allow comparison of the results within any set, the m.s.e. at $K = 0$ and $a = 0$ is normalized to unity. The ratio, D , of the value of the minimum m.s.e. at any (K, a) to the value of m.s.e. at $K = a = 0$, is used to compare different sets of results. All systems have a pole at the origin, unless otherwise stated.

Since the second-order system behaves rather differently from higher order systems, it is dealt with as a separate case. Figures 3(a) and (b) detail the



(a) Variation with the forward gain, for a second-order system. Rate gain $a = 0$.



(b) Variation with rate feedback gain, a , for a second-order system with $N = 2, x = 2$.

Fig. 3. Variation of mean square error.

results. The salient points are that such a system cannot be optimized using forward gain, since no minima exist; while optimization is possible using rate gain, the curves possess such sharp minima (for

worthwhile error reduction) and are so asymmetric that in practice it would be difficult to maintain the minimum value. The second-order system with a first-order filter pole is capable of analytic evaluation using the methods of Solodovnikov,⁹ and yields the result:

$$\begin{aligned} \bar{\epsilon}^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{\theta_i(j\omega_1)}{1+G(j\omega_1)} \right|^2 d\omega_1 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega_1}{|(j\omega_1+N)^x| |1+G(j\omega_1)|^2} \\ &= \frac{N^2}{4} + \frac{(N^2-N+K)(N^2-1)}{4\left(1-2K-N^2-\frac{K^2}{N^2}\right)} \quad x = 1. \end{aligned}$$

Differentiation with respect to K indicates that no minima exist, and the calculated curves from the expression are shown dotted in Fig. 3(a).

Third- and higher-order systems produce curves of the general form shown in Fig. 4, with variation of K ($a = 0$). Here, for worthwhile error reduction, the curves tend to flat minima, and the filter pole must be much closer to the origin than the principal pole of the open-loop system (i.e. $N < 1$). The comparison of different orders of system, against position of the

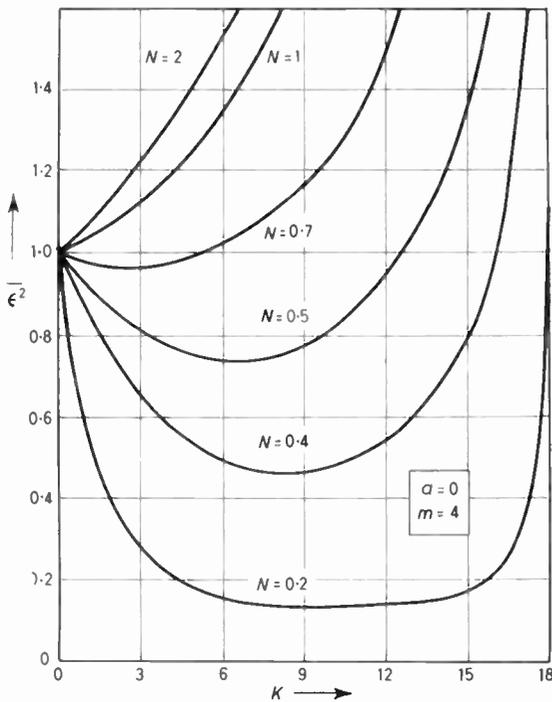


Fig. 4. General form of variation of mean square error with forward gain, $x \leq 3$.

filter pole, is shown in Fig. 5. Also shown on this figure are the values of $s = S_b$ at which the stability-determining or major, branch of the system root-locus breaks away from the negative real axis. It was noted that for values of $N \leq S_b$, D is related to N by $D = c(N)^m$, for the particular systems studied. In this, c is a linear function of the order of the system. Since the pole spacing was arbitrarily chosen, it might be concluded in general that altering the position of the filter pole will have much greater effect upon the value of the minimum m.s.e. than changing the order of the system, say, by compensating networks.

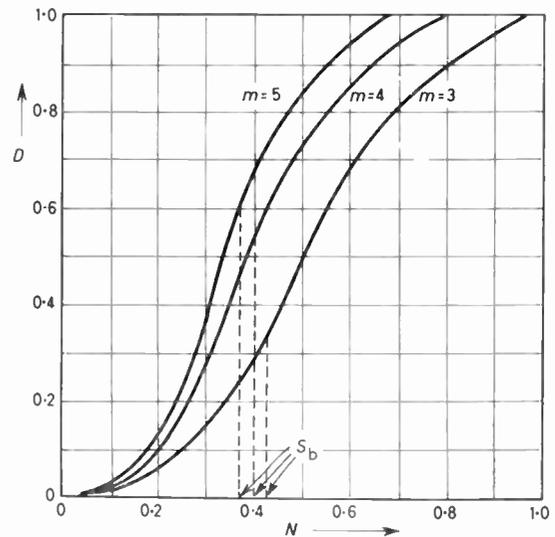


Fig. 5. A comparison of the performance of optimization of systems of increasing order, for $x = 2$.

The general effect of including a zero in the forward transfer function is shown in Fig. 6. For any given order of system and fixed filter-pole position, as the zero is steadily moved inwards towards the origin on the s -plane the break-away point (S_b) of the root-locus will move outwards, i.e. the value of the filter-pole decreases relative to the break-away point. Thus from Fig. 6, it would appear that it is the value of N relative to S_b in any system that dictates the depth of minimum of m.s.e.; N must be very much less than S_b for worthwhile reductions in m.s.e. with forward gain. The rise in D in Fig. 6, when the zero approaches the origin, is consistent with results described later for systems with no pole at the origin.

The use of rate feedback gain as an optimizing parameter requires some care. First of all, the use of rate feedback distorts the shape of the m.s.e. to forward-gain curves, Fig. 7, but greatly reduces the m.s.e. The conditional stability with a is well demon-

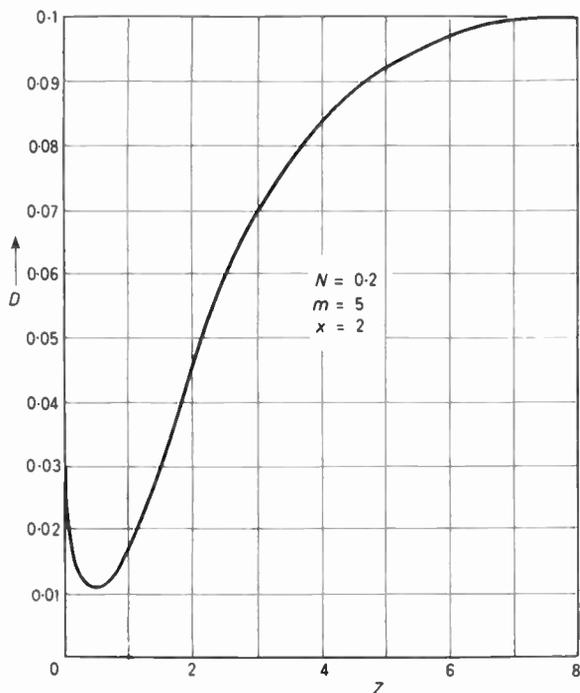


Fig. 6. The effect of varying a zero in the function $F(s_1)$, for $N = 0.2, m = 5, x = 2$.

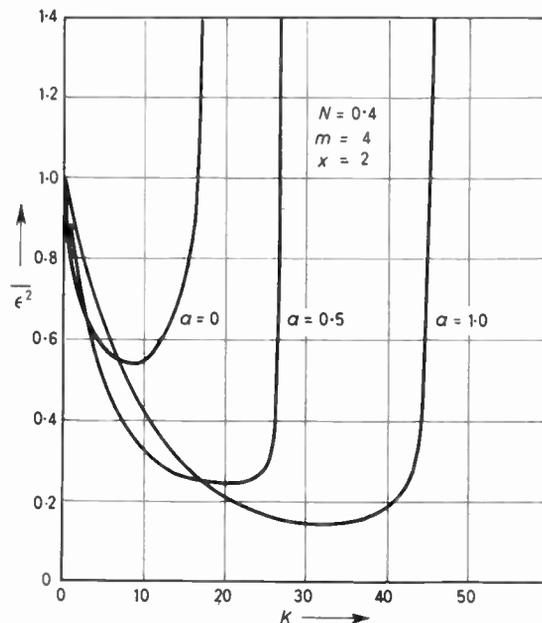


Fig. 7. Distortion due to rate feedback in a fourth-order system when $N = 0.4, x = 2$.

strated in Fig. 8, as is also the fact that when a is optimized there will also be little point in increasing K beyond some value (about 40 in this case). In other words, two-parameter optimization is possible. The real problem, though, is typified by Fig. 9, for a given system with given forward gain, in which the m.s.e. is minimized by varying a . As the filter pole moves towards the origin, the minima decrease in depth and occur at ever reducing values of a . Finally, for low values of N the minima disappear altogether. This is in total contradiction to the behaviour of depth of minima with the value of N for optimization with K (cf. Figs. 5 and 4).

Therefore, it appears that if a system is to be optimized using K or a , or both, different criteria apply to the position of the equivalent noise filter pole N , relative to the root-locus break-away. If K alone is used, then N should be much nearer the origin than the break-away. Alternatively, if only a is being used to optimize the system, then N should be at or to the left of the break-away for good minima; the farther N is to the left, the higher the m.s.e. at the minima, but the more pronounced are the minima. Thus for two parameter optimization of the systems studied, a reasonable compromise would be positioning N at, or slightly to the right (nearer origin) of, the break-away of the major branch of the root-locus.

Systems with no pole at the origin (i.e. possessing finite static error) follow the above findings in a broad sense, the values of the actual m.s.e. naturally being different. When the principal pole of such a

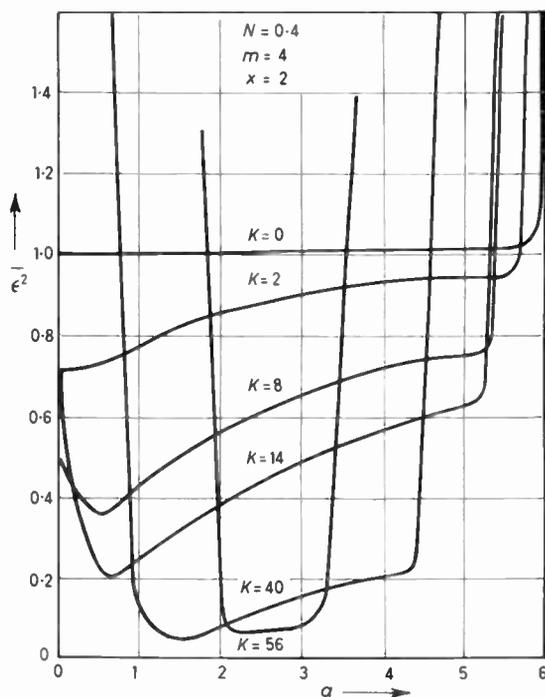


Fig. 8. The effect of rate feedback on higher-order systems, illustrated by a fourth-order system when $N = 0.4, x = 2$.

system is taken as the origin on the s -plane, then the comments of the preceding paragraph regarding the positioning of the filter pole are still valid. However, the curves of m.s.e. to K or a are highly asymmetric, and Fig. 10 demonstrates this using K as the parameter. Because of this, and the obvious practical limitation of such systems, detailed results are not presented, nor would such systems be particularly suitable for optimization procedure.

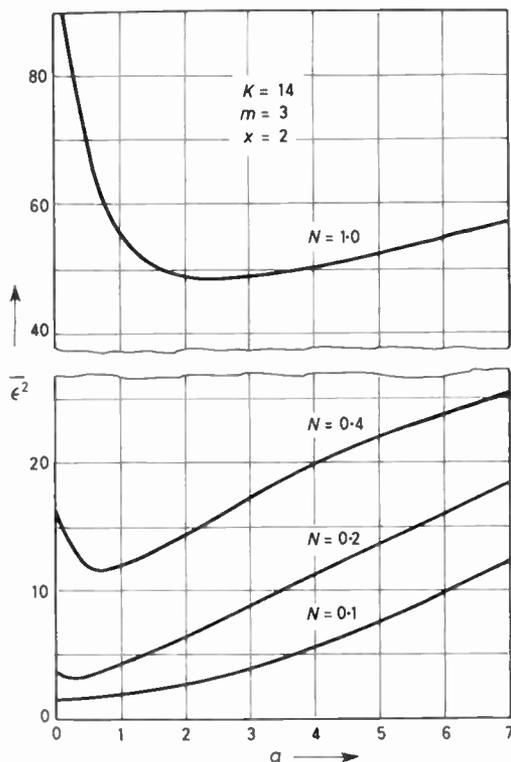


Fig. 9. The disappearance of minima as the filter pole approaches the origin, for $m = 3, K = 14, x = 2$.

Finally, the effect of increasing x , the order of the filter pole, is basically to increase the asymmetry of all curves, particularly for low values of N . The values of N at which minima are first observed increases slightly as x increases, and therefore the depth of minima in a given system also increases with x . Generally, for the system studied, the asymmetry was very noticeable for $x = 3$, and for $x = 4$ this became excessive, with very sharp minima. It would appear that if the frequency spectrum of the random disturbance is represented by a third or higher order pole, difficulty may be encountered in the stability of the optimizing loops.

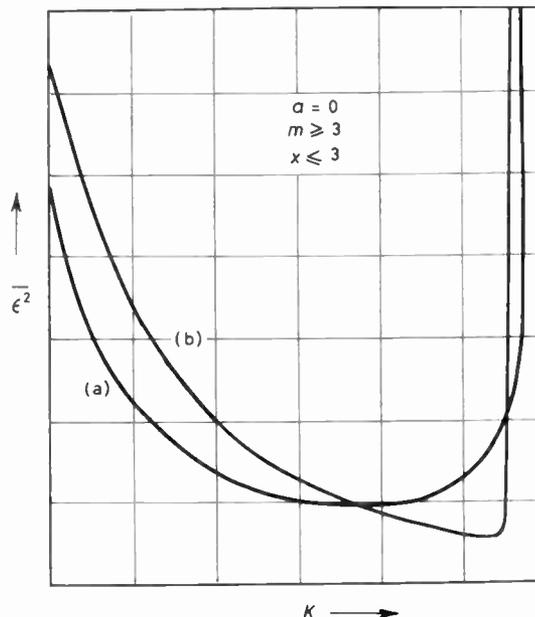


Fig. 10. General comparison of systems (a) with integration, and (b) with no integration, in the loop, for $a = 0, m \geq 3, x \leq 3$

5. Conclusion

The experimental study was limited to systems with real poles and zeros in the forward transfer function. The choice of pole spacing was arbitrary, and contributes no peculiarities to the mathematical functions. Multiple poles were not considered. From the consistency of the trend of the results for all orders of systems considered, it is thought that generalizations may be made. Thus, if optimization is to be carried out using the forward- and rate-stabilizing gains, the forward transfer function must be so designed that the x -th order pole characterizing the disturbing signal lies in the near vicinity of the break-away of the major branch of the root-locus. The system designer, therefore, still requires considerable prior knowledge. Finally the assumption that is so often made in theoretical analyses of parabolic curves of mean square error to optimizing parameter has been shown to be invalid for the majority of the cases investigated. The shapes of the curves suggest that more detailed attention should be given to the optimizing loops themselves, to prevent pointless roaming over very flat minima, and possible instability¹⁰ either at sharp minima or at very sharp transitions from flat minima to steeply rising sides.

6. References

1. P. H. Hammond and M. J. Duckenfield, 'Automatic optimization by continuous perturbation of parameters', *Automatica*, 1, No. 2-3, p. 147, August 1963.

2. K. C. Ng, 'High frequency perturbation in hill-climbing systems', *Proc. Instn Elect. Engrs*, 111, p. 1907, November 1964.
3. J. D. Roberts, 'Extremum or hill-climbing regulation' *Proc. I.E.E.*, 112, p. 138, January 1965.
4. R. S. Phillips, 'Theory of Servomechanisms' (McGraw-Hill, New York, 1947).
5. V. V. Solodovnikov, 'Introduction to the Statistical Dynamics of Automatic Control Systems', p. 164 (Dover Publications, New York, 1960).
6. C. H. Wilts, 'Principles of Feedback Control', p. 132 (Addison-Wesley, London, 1960).
7. O. L. R. Jacobs, 'Measuring mean square value of random signals', *J. Electronics and Control*, 9, p. 149, 1960.
8. Z. J. Jelonek, A. B. Gardiner and D. Raeside, 'A theoretical comparison of three types of self-optimizing control systems', Joint Instn Mech. Engrs-U.K. Automation Council Convention on Advances in Automatic Control, Nottingham, 5th-9th April 1965. (Paper No. 12).
9. V. V. Solodovnikov, *ibid*, p. 124 and p. 299.
10. Z. J. Jelonek, P. L. V. Pomella and N. S. Karunaratne, 'Oscillations in feedback systems subjected to periodic interference, and stability of a self-optimising control system', *Proc. I.E.E.*, 111, p. 1881, November 1964.

Manuscript first received by the Institution on 4th July 1966 and in final form on 28th December 1966. (Paper No.1104/C93.)

© The Institution of Electronic and Radio Engineers, 1967

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10} , from nominal frequency for February 1967

February 1967	24-hour mean centred on 0300 U.T.			February 1967	24-hour mean centred on 0300 U.T.		
	GBZ 19.6 kHz	MSF 60 kHz	Droitwich 200 kHz		GBZ 19.6 kHz	MSF 60 kHz	Droitwich 200 kHz
1	—	—	— 0.7	15	— 299.2	— 300.2	+ 0.9
2	—	— 299.9	— 0.6	16	— 298.0	— 300.5	+ 0.1
3	— 301.4	— 299.7	— 0.3	17	— 300.2	— 301.2	+ 0.1
4	— 300.8	— 298.8	+ 0.3	18	— 300.9	— 297.7	0
5	—	— 299.9	+ 1.0	19	— 300.1	— 298.5	— 0.1
6	— 302.0	— 299.2	+ 1.3	20	—	— 297.8	0
7	— 301.9	— 299.6	+ 1.2	21	—	— 298.7	— 0.2
8	— 301.2	— 299.3	+ 1.2	22	— 301.3	— 298.6	— 0.5
9	— 301.9	— 298.4	+ 1.2	23	— 300.9	— 300.3	— 0.5
10	— 300.7	— 300.1	+ 0.5	24	— 302.0	— 299.3	— 0.5
11	— 300.3	— 300.8	+ 0.6	25	— 298.9	— 299.7	— 0.8
12	— 300.8	— 300.6	+ 0.5	26	— 297.2	— 300.6	— 1.4
13	— 301.5	— 300.7	+ 0.8	27	— 299.1	— 299.9	— 1.1
14	— 301.9	— 300.6	+ 0.6	28	— 297.5	— 300.1	— 1.6

Nominal frequency corresponds to a value of 9 192 631 770.0 Hz for the caesium $F_{m(4,0)}-F_{m(3,0)}$ transition at zero field.
Mean monthly values: GBZ — 300.6, MSF — 299.6

Changes to the MSF Standard Frequency Transmissions

As from the 1st April 1967 it is proposed to control the MSF Standard frequency transmissions from a rubidium gas cell frequency Standard, and this will coincide with a change in policy regarding the maintenance of the carrier frequencies and the modulation time-signals on a coherent basis.

From this date the carrier frequencies of the MSF transmissions will be operated, without frequency offset, at nominal frequency in the atomic scale.

The MSF modulation time-signals, and also the carrier frequency and time-signals of the GBR 16 kHz transmission will continue to be offset from nominal frequency by the amount internationally agreed upon at the beginning of each year.

Inductorless Band-pass I.F. Amplifiers

By

A. WORONCOW, Dipl. Ing.†

AND

J. CRONEY, B.Sc., Ph.D.†

Summary: RC interstage coupling filters can be used in a cascaded transistor amplifier to produce a gain of 77 dB at a mid-frequency of 60 MHz with 20 MHz bandwidth. A logarithmic amplifier can be arranged by using active band-pass filters of unity gain in conjunction with the RC coupled stages.

1. Introduction

In a previous paper¹ the authors described a new type of logarithmic i.f. amplifier, and have subsequently been concerned with its development in an integrated circuit form. However the inclusion of a large number of band-pass coupled inductors detracts from the full advantage of an integrated circuit lay-out. The present paper describes work aimed at obtaining interstage coupling circuits using resistance-capacitance elements which give very similar band-pass performance to the usual inductance-capacitance arrangement.

The possibilities of coil-less filters have been realized for some time, but they have received special attention recently because the high current gains of transistors make them a practical proposition. Numerous circuits suitable for application at audio frequencies, and at radio frequencies of a few hundred kilohertz, have been published in the literature (for example, references 2 to 6). In general these circuits are rather complex and are therefore difficult to apply to radar i.f. amplifiers, which are usually required to work at some tens of megahertz. In the present paper a circuit is described which can work at frequencies up to 100 MHz and which is especially suitable for wide band-pass amplifiers such as are used with short pulse radar.

2. Resistance-Capacitance Band-pass Filters

The simplest resistance-capacitance band-pass filter is formed by a resistance-capacitance low-pass filter followed by a resistance-capacitance high-pass filter as shown in Fig. 1. This kind of resistance-capacitance band-pass filter has high attenuation and has a large bandwidth relative to its mid-frequency. Nevertheless, it is possible to build an amplifier using such filters because the gain obtainable with transistors is sufficiently high. The selectivity, however is poor and makes such an amplifier in general unacceptable. Also with present-day transistors it is difficult to obtain useful gain above 30 MHz.

The Q of such a band-pass circuit may be increased, and also more gain obtained at the higher frequencies when the band-pass filter (Fig. 1(c)) is applied as a positive feedback loop in an amplifier. A very suitable

form of amplifier for this purpose is a long-tailed pair, such as shown in Fig. 2, where R2 and C2 form the low-pass filter, and C3 and R5 form the high-pass filter. The composite band-pass filter produces zero phase shift at the 'resonant' frequency, so at this frequency there is positive feedback from the base of TR2 round the loop comprising R7 and R3 (R4 is assumed to be of much larger value than these resistors) and the filter network; at other frequencies, because of the attenuation and phase shift of the filter network, the overall gain is greatly reduced. The inclusion of the resistors R3 and R7 is in itself degenerative, and if their values are too small the circuit forms a sine-wave oscillator. Thus R3 and R7 control the current gain at the 'resonant' frequency; without affecting these parameters the voltage gain of the amplifier can be varied by adjusting the value of R6.

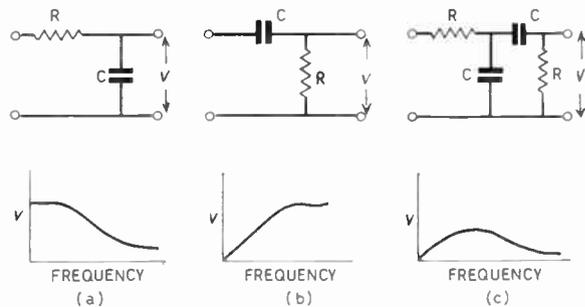


Fig. 1. Low-pass, high-pass and band-pass RC filters.

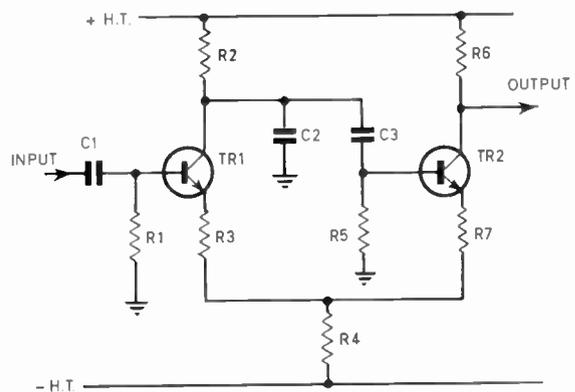


Fig. 2. Basic circuit of RC active band-pass filter using a long-tailed pair.

† Admiralty Surface Weapons Establishment Extension, Funtington, near Chichester, Sussex.

The following advantages of the circuit should be noted. The input terminal is free and the impedance of the feeding circuit has little effect on the shape of the frequency characteristic. Likewise, the output terminal is free and the load has little effect on the feedback loop. The circuit is extremely simple and the positive feedback reinforces the amplification of the transistors thus making it possible to apply the circuit at some tens of megahertz. Above 100 MHz, however, the capacitances and resistances of the electrodes, and the phase shifts inside the transistors, make the circuit highly degenerative.

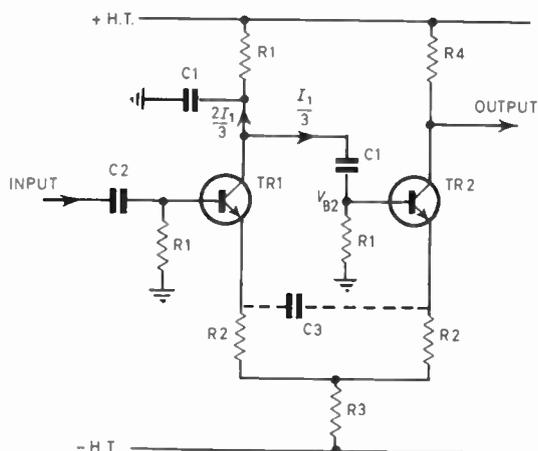


Fig. 3. Amplifier with equal R and C in the band-pass filter.

A variant of the circuit of practical importance and which also makes its operation easier to understand, is shown in Fig. 3. The resistors, R_1 , and capacitors, C_1 , in the band-pass filter are equal. The 'resonant' frequency occurs when the impedance of C_1 is equal to the resistance of R_1 and under these conditions the circuit produces maximum phase shift with change of frequency. If an alternating current I_1 is fed into the filter network at the 'resonant' frequency with the loop open at the collector of TR1, current $I_1/3$ flows in the base resistor of TR2 (Fig. 3) and a current in '2R2' of $I_1/3$ times $R_1/2R_2$, assuming R_3 and the current gains infinite. When the loop is closed, I_1 is the collector current of TR1 and for less than unity loop gain $I_1 R_1/6R_2 < I_1$. Thus if $R_2 > R_1/6$, the circuit is unconditionally stable, because the current gain round the positive feedback loop is less than unity however large the current gain of the transistors. Since there is some internal resistance in the emitters and the bases of the transistors, and also a limit to their current gain, the critical value of R_2 is somewhat smaller, but the above reasoning enables one to make an approximate estimate for a start. It may be seen that the 'resonant' frequency occurs when there is no phase shift between the collector current I_1 , and the voltage V_{B2} at the base of the transistor TR2. At the collector of TR1, the voltage is lagging by 45°

with respect to the current I_1 and this phase shift is corrected by the high-pass filter. Unequal values of capacitances or resistances may produce phase shifts other than 45° but the basic property remains, that the lagging shift is cancelled by the leading shift at the 'resonant' frequency.

At higher frequencies there is a lagging phase shift in the transistors between the signal voltage on the base, and the emitter or collector current. Above about 30 MHz, it becomes necessary to compensate this phase shift even with the fastest transistors available at present, otherwise very noticeable deviation of frequency occurs. This can be done very conveniently by means of the capacitor C_3 shown in Fig. 3, but it has to be taken into account that such a capacitor lowers the impedance in the emitters, and raises the gain. Although, theoretically, a Q of any value is obtainable with the circuit, it can be seen that high Q exists when the circuit is near to oscillation point. Thus the high- Q circuit depends very strongly on stability of components and transistors. It seems, therefore, that the circuit in this simple form is suitable for relatively large bandwidth application, where a Q of not greater than about 10-15 is required. Improvement of selectivity would probably be easier to achieve by increasing the number of stages in the amplifier, than by raising the circuit Q .

3. Amplifier Circuit

An i.f. amplifier at 60 MHz with 20 MHz bandwidth has been built with resistance-capacitance band-pass filters and employing resistance-capacitance coupling between the stages. There are six stages, staggered to give a flat-top response, and the circuit-diagram of two staggered stages is shown in Fig. 4. The overall gain of the six stages is 77 dB, and the frequency

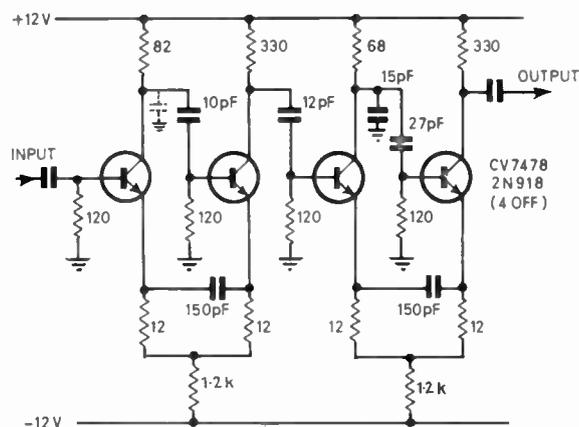


Fig. 4. Staggered pair of amplifiers with RC active filters. Intermediate frequency 60 MHz. Bandwidth 20 MHz.

Tail resistors of 1.2 kΩ give a standing current of about 5.0 mA. Collector capacitance of the 1st transistor is the shunt capacitance of the filter.

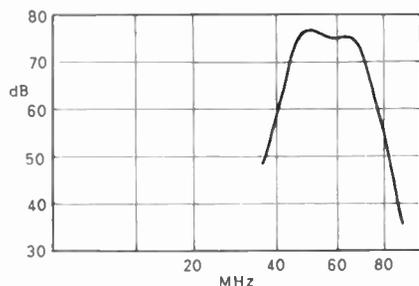


Fig. 5. Response characteristic of RC band-pass Amplifier 6 stages—active filters.

response is shown in Fig. 5. The resistance-capacitance coupling between stages has a very short time-constant in order to minimize the possibility of paralysis by strong signals, and other time-constants of the circuit are also very short. The values in the band-pass filters should not be regarded as the only ones possible for this amplifier as a similar response can be obtained by other combinations of capacitances and resistances. The amplifier has been tested on appropriate signal generators and its behaviour was remarkably similar to that of amplifiers constructed with traditional tuned-inductance circuits. There was no paralysis after even the strongest short pulses, and there were no particular saturation or overload effects. The reduction of such defects is helped by the peculiar limiting properties of long-tailed pairs.

Seeing that the overall gain of a stage with a resistance-capacitance filter can be controlled independently of the band-pass property, the circuit can be utilized as an active band-pass filter with a gain of unity. This should make it possible to construct, without inductance-tuned circuits, a logarithmic amplifier without successive detection (e.g. a true i.f. logarithmic amplifier).¹ The general scheme is shown in Fig. 6. The filter has a gain of unity and should accommodate without limiting the largest possible signal coming out of the two amplifiers in parallel. It is thought that five or six such stages should make a complete logarithmic amplifier with a dynamic range of 80–90 dB and giving good approximation to the ideal logarithmic law. Since the values of capacitances

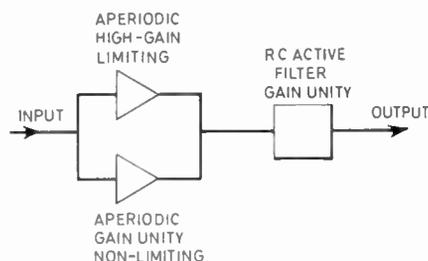


Fig. 6. A stage of true i.f. logarithmic amplifiers with RC circuit.

and resistances in Fig. 4 are in the range most suitable for integrated circuits, it should be possible to make 60 MHz amplifiers, whether linear or logarithmic, in the form of relatively simple integrated circuits.

4. Acknowledgments

The authors thank Mr. C. Street for assistance in the experimental work and the Ministry of Defence (Navy) for permission to publish the paper.

5. References

1. A. Woroncow and J. Croney, 'A true i.f. logarithmic amplifier using twin-gain stages', *The Radio and Electronic Engineer*, **32**, No. 3, pp. 149–55, September 1966.
2. J. G. Linvill, 'Resistance-capacitance active filters', *Proc. Inst. Radio Engrs*, **42**, pp. 555–64, March 1954.
3. A. A. Gaash, 'Synthesis of Integrated Selective Amplifiers for Specified Response and Desensitivity', Electronics Research Laboratory, University of California, Berkeley. E.R.L. Report No. 65–31.
4. L. Scott, 'Criteria for the design of active filters using resistance and capacitance elements in feed-back circuits', *Solid State Electronics*, **9**, pp. 641–51, 1966.
5. D. A. Calahan, 'Sensitivity minimization in active RC synthesis', *Trans. Inst. Radio Engrs on Circuit Theory*, **CT-9**, No. 1, pp. 38–42, March 1962.
6. S. S. Hakim, 'RC active filters using an amplifier as the active element', *Proc. Instn. Elect. Engrs*, **112**, No. 5, pp. 901–12, May 1965.

Manuscript received by the Institution on 10th February 1967. (Contribution No. 96.)

© The Institution of Electronic and Radio Engineers, 1967

The Response of Mismatched Prototype Image Impedance Filters

By

N. H. BRIGHT,
C.Eng., A.M.I.E.R.E.†

Summary: A normalized low-pass prototype filter, terminated in equal value resistances, is considered and the insertion loss is given, as a function of normalized frequency and value of the terminating resistance. The root-locus of the poles of the insertion-loss function is plotted showing the effect of variation of the value of the termination and comparison is made with the pole locations of Butterworth and Chebyshev filters of the third order. The treatment is extended to two identical prototype filters in cascade and in this case the root-locus of the poles of the insertion-loss function is compared with the poles of fifth order Butterworth and Chebyshev filter functions. Insertion-loss curves are given for various terminations for both single and cascaded filters.

1. Introduction

This paper presents an attempt to link image-parameter filters to modern insertion-loss theory. The first systematic methods of synthesizing two-port networks were based on the theory of image-parameters. An extensive and adequate literature exists, see, for example, Zobel¹ and Guillemin.² The major drawback of filters designed on an image-parameter basis is that the ideal attenuation and phase characteristics are not, in practice, obtained. The ideal theoretical characteristics relate to the condition of correct termination where the filter is terminated in its image impedances at all frequencies of significance. Since, however, the image impedances are functions of frequency, then correct termination is difficult to achieve. In practice the filter is terminated in fixed value resistances normally made equal to the design impedance, R_0 , of the filter and this gives rise to a response which is different to that of correct termination. In this paper the effect of termination in resistances which are not equal to the design impedance is considered.

2. The Normalized Prototype Filter-section

Figure 1 shows the T and π prototype low-pass filter-sections. It is shown by Frazer³ that the design impedance, R_0 , and the angular cut-off frequency, ω_{c0} , of the sections shown are given by

$$R_0 = \sqrt{\frac{L}{C}} \quad \dots\dots(1)$$

and

$$\omega_{c0} = \frac{2}{\sqrt{LC}} \quad \dots\dots(2)$$

From equations (1) and (2)

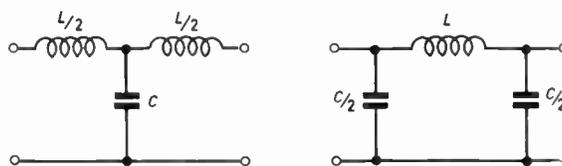


Fig. 1. T and π -prototype low-pass filters.

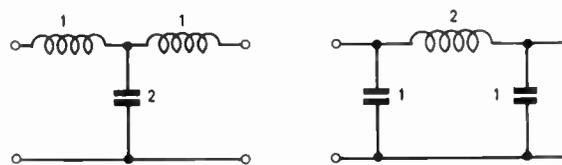


Fig. 2. Two basic 'model' low-pass filters.

$$L = \frac{2R_0}{\omega_{c0}} \quad \dots\dots(3)$$

and

$$C = \frac{2}{\omega_{c0}R_0} \quad \dots\dots(4)$$

If now R_0 is normalized to a convenient value of 1 ohm and ω_{c0} is reduced to 1 rad/s, then

$$L = 2 \text{ henrys}$$

and

$$C = 2 \text{ farads}$$

Thus the two basic 'model' low-pass filters are obtained, both having $R_0 = 1$ ohm and $\omega_{c0} = 1$ rad/s. These are shown in Fig. 2.

2.1. The Insertion Loss of a Single Prototype Filter

Referring to Fig. 3 the insertion loss, in decibels, is defined as

$$20 \log_{10} |I_1/I_2|$$

In the case of the T-section the terminations are taken

† Department of Electrical Engineering, Wolverhampton College of Technology, Wolverhampton.

as resistances of n ohms, whereas for the π -sections the terminations are n_1 ohms. By standard circuit analysis it may be shown that, for the T-section

$$\frac{I_1}{I_2}(s) = 1 + s\left(\frac{1+n^2}{n}\right) + 2s^2 + \frac{s^3}{n} \dots\dots(5)$$

and

$$\left|\frac{I_1}{I_2}(x)\right|^2 = (1-2x^2)^2 + x^2\left(\frac{1+n^2}{n} - \frac{x^2}{n}\right)^2 \dots\dots(6)$$

Similarly, for the π -section,

$$\frac{I_1}{I_2}(s) = 1 + s\left(\frac{1+n_1^2}{n_1}\right) + 2s^2 + n_1s^3 \dots\dots(7)$$

and

$$\left|\frac{I_1}{I_2}(x)\right|^2 = (1-2x^2)^2 + x^2\left(\frac{1+n_1^2}{n_1} - n_1x^2\right)^2 \dots\dots(8)$$

where s = the complex frequency variable, and $x = \omega/\omega_{c0}$, the normalized frequency.

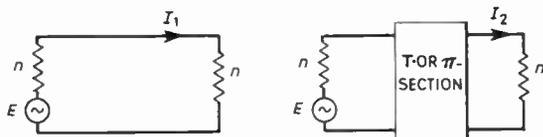


Fig. 3. Circuits used to define the insertion loss of a network.

Now eqns. (6) and (8) become identical if $n = 1/n_1$, and hence a function is obtained which relates insertion loss to the frequency, x , with n (or n_1) as the parameter which represents variation of termination.

2.2. Pole Locations

The network function for the T-section is the reciprocal of eqn. (5):

$$\frac{I_2}{I_1}(s) = \frac{1}{1 + s\left(\frac{1+n^2}{n}\right) + 2s^2 + \frac{s^3}{n}} \dots\dots(9)$$

and for the π -section, the network function is the reciprocal of eqn. (7):

$$\frac{I_2}{I_1}(s) = \frac{1}{1 + s\left(\frac{1+n_1^2}{n_1}\right) + 2s^2 + n_1s^3} \dots\dots(10)$$

Equations (9) and (10) are third-order 'all pole' functions and their pole locations will depend on the value of the terminations. The pole locations were determined with the aid of a digital computer and are given in Table 1.

The migration of the poles of the T-section, as the termination value is varied, is shown plotted in Fig. 4,

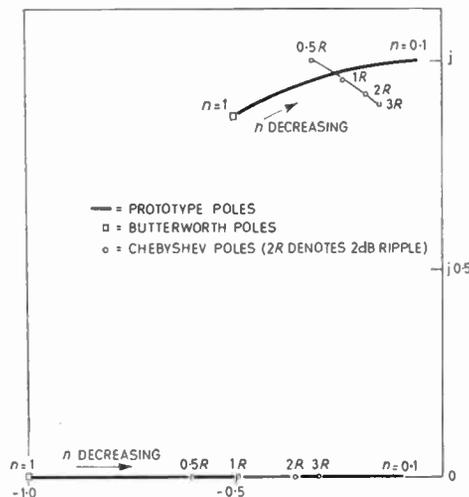


Fig. 4. Root-locus showing migration of poles of a single prototype filter as the value of the termination is varied. (Upper half only of complex plane shown.)

Table 1
Pole locations for single section

n	n_1	Pole locations
1.0	1.0	$-0.50 \pm j0.866$ -1.0
0.9	1.11	$-0.45 \pm j0.893$ -0.90
0.8	1.25	$-0.40 \pm j0.9165$ -0.80
0.7	1.428	$-0.35 \pm j0.9367$ -0.700
0.6	1.667	$-0.30 \pm j0.9539$ -0.60
0.5	2.00	$-0.25 \pm j0.9682$ -0.50
0.4	2.50	$-0.20 \pm j0.9798$ -0.40
0.3	3.33	$-0.15 \pm j0.9887$ -0.30
0.2	5.0	$-0.10 \pm j0.9950$ -0.20

and this shows also the pole locations for Butterworth and Chebyshev third-order filters. In the case of the Chebyshev filter the poles are given for different values of pass-band ripple (in dB): these were taken from Weinberg.⁵ It may be seen from Fig. 4 that as n is varied from unity towards zero, the poles of the prototype filter move towards the 'j ω ' axis and pass through the region of the Chebyshev poles. Hence an approximation to the Chebyshev equi-ripple behaviour in the pass-band may be expected for some

fractional values of n . Also, when $n = 1$, the poles of the prototype filter coincide with the third-order Butterworth poles and thus $n = 1$ gives the condition of maximal flatness in the pass-band. Hence, for the T-section, a reduction in the value of n from unity will result in a transition from the maximally flat characteristic towards the ripple characteristic of the Chebyshev filter and obviously the magnitude of the ripple will increase as n is decreased. For the π -section, an increase in the value of n_1 will have exactly the same effect on the poles of the filter as a decrease in n had for the T-section. Thus, if $n = 0.5$, for the T-section and $n_1 = 2$, for the π -section, then the characteristics of the two filters will be identical. It may be seen, for example, from Fig. 4 and Table 1, that for $n = 0.5$ the poles are quite close to the Chebyshev poles, for 1 dB ripple, and so a ripple of the order of 1 dB is to be expected for this value of termination.

3. Insertion Loss for Single Section

The insertion loss, in decibels, for the T-section, was calculated from eqn. (6) over a wide frequency range and for different values of n and a selection of the results are shown plotted in Fig. 5. At frequencies well above cut-off ($x \gg 1$) the rate of attenuation is seen to be asymptotic to a slope of 18 dB/octave as would be expected for a three-pole filter. The attenuation rate in the stop-band, immediately adjacent to the cut-off frequency, is seen to vary considerably with the termination value. For many applications, where pass-band ripple can be tolerated, then values of n less than unity may well be desirable in order that the steeper cut-off may be obtained. The variation of

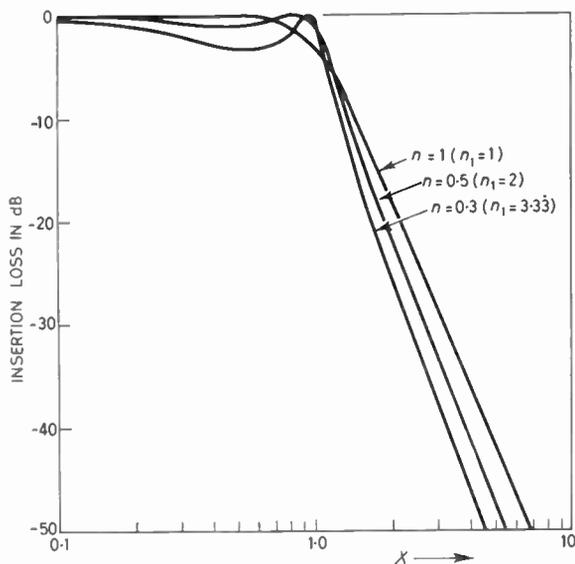


Fig. 5. Insertion loss of normalized prototype filters.

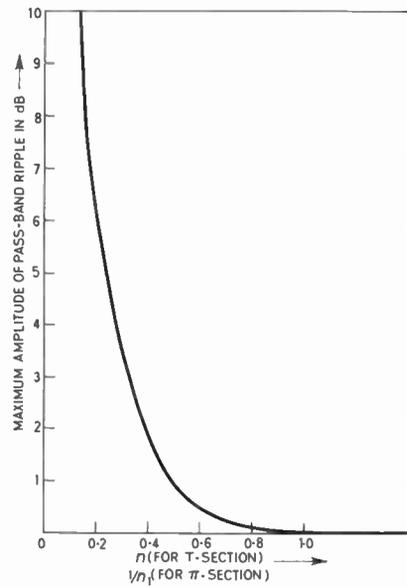


Fig. 6. Approximate value of pass-band ripple as a function of n or n_1 for a single prototype section.

pass-band ripple as a function of termination is plotted in Fig. 6. This shows that, for the T-section, the range, $0.3 \leq n \leq 1$, is most likely to be of practical value. Similarly, for the π -section, the range, $1 \leq n_1 \leq 3.33$, is of main interest.

3.1. The Insertion Loss of Two Identical Prototype Filters in Cascade

In Fig. 7, F_1 and F_2 are identical normalized sections, either both T- or both π -sections. (If π -sections are used the terminations will be n_1 .) Analysis leads to the following equations:

$$\frac{I_1}{I_2}(s) = 1 + 2s \left(\frac{1+n^2}{n} \right) + 8s^2 + 2s^3 \left(\frac{3+2n^2}{n} \right) + 8s^4 + \frac{4s^5}{n} \dots(11)$$

and

$$\left| \frac{I_1}{I_2}(x) \right|^2 = [1 - 8x^2 + 8x^4]^2 + 4x^2 \left[\frac{1+n^2}{n} - x^2 \left(\frac{3+2n^2}{n} \right) + \frac{2x^4}{n} \right]^2 \dots(12)$$

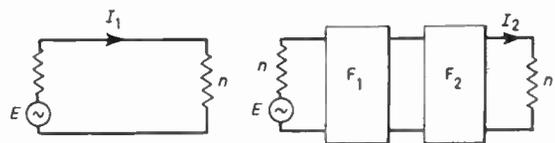


Fig. 7. Two identical prototype filters in cascade.

for the T-sections, and to:

$$\frac{I_1}{I_2}(s) = 1 + 2s \left(\frac{1+n_1^2}{n_1} \right) + 8s^2 + 2s^3 \left(\frac{2+3n_1^2}{n_1} \right) + 8s^4 + 4n_1s^5 \dots\dots(13)$$

and

$$\left| \frac{I_1}{I_2}(x) \right|^2 = [1 - 8x^2 + 8x^4]^2 + 4x^2 \left[\frac{1+n_1^2}{n_1} - x^2 \left(\frac{2+3n_1^2}{n_1} \right) + 2n_1x^4 \right]^2 \dots\dots(14)$$

for the π -sections. Equations (12) and (14) become identical if $n = 1/n_1$.

3.2. Pole Locations

The network functions for the T- and π -sections are the reciprocals of eqns. (11) and (13) respectively and both are 'all pole' fifth-order functions. The pole locations are given in Table 2. The migration of the poles of the network function is shown in Fig. 8 and also shown are the poles of fifth-order Chebyshev filters.

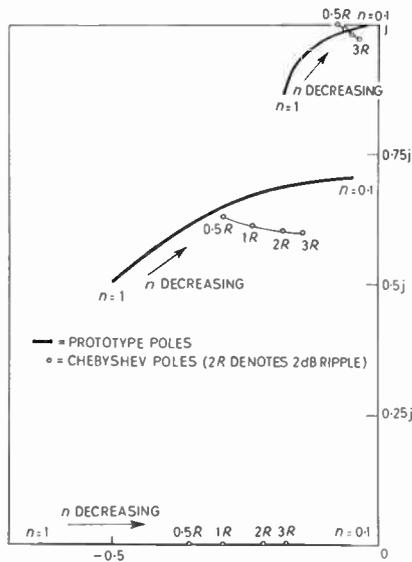


Fig. 8. Root-locus showing migration of poles of two prototype filters in cascade. (Upper half only of complex plane shown.)

For $n = 1$ the poles of the cascaded sections are nearest to being located on the circumference of a circle centred on the origin and hence this termination condition would be expected to approximate to the maximally flat response of the Butterworth filter. However, the radius of the circle passing through the mean location of the poles is much less than unity

and this suggests that the cut-off frequency (defined as where the insertion loss is 3 dB not including ripple) will be less than 1 rad/s.

Table 2 Pole locations for two cascaded sections

n	n ₁	Pole locations	
1.00	1.00	-0.6478	-0.50 ± j0.50 -0.176 ± j0.861
0.90	1.11	-0.556	-0.45 ± j0.545 -0.172 ± j0.883
0.80	1.25	-0.473	-0.40 ± j0.583 -0.163 ± j0.905
0.70	1.428	-0.398	-0.35 ± j0.614 -0.151 ± j0.926
0.60	1.667	-0.329	-0.30 ± j0.640 -0.135 ± j0.945
0.50	2.00	-0.267	-0.25 ± j0.661 -0.117 ± j0.961
0.40	2.50	-0.208	-0.20 ± j0.678 -0.0978 ± j0.975
0.30	3.333	-0.153	-0.15 ± j0.691 -0.075 ± j0.986
0.20	5.00	-0.100	-0.100 ± j0.700 -0.050 ± j0.994
0.10	10.00	-0.05	-0.050 ± j0.705 -0.025 ± j0.998

Further, since the poles are not exactly equally spaced on the circumference of a circle then the response of the cascaded sections will not be quite flat in the pass-band but may be expected to show a small amount of ripple. As n is decreased the poles of the cascaded sections are seen to move towards the $j\omega$ axis and to approach the pole locations of the Chebyshev fifth-order filters. Hence, as n is decreased from unity an increasing amount of pass-band ripple is to be expected.

3.3. Insertion Loss of Two Cascaded Sections

The insertion loss was computed from eqn. (12) over a wide range of frequency and for various values of n . A selection of the results obtained are shown plotted in Fig. 9. At frequencies well above cut-off the rate of attenuation becomes asymptotic to 30 dB/octave as would be expected for a five-pole filter. The same general principles apply as for the single section, namely, that the greater the degree of mismatch allowed at the terminations the larger becomes the magnitude of the ripple. For most values of n , a secondary ripple, of much smaller magnitude than the main ripple, is introduced into the pass-band. However, for practical tolerable values of main ripple the secondary ripple is almost negligible. Thus, for $n = 0.4$, the main ripple is about 2.7 dB, whereas the

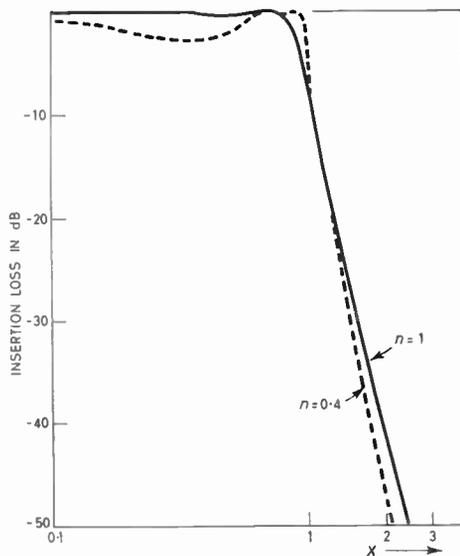


Fig. 9. Insertion loss of two prototype sections in cascade.

secondary ripple is around 0.2 dB. Figure 10 shows the variation of ripple as a function of termination; the range $0.4 \leq n \leq 1$ would seem to be of the main practical value.

4. Conclusion

An attempt has been made to relate classical filters to modern filter theory. It has been shown that the Chebyshev equi-ripple characteristic is obtainable by termination of prototype filters in a resistance not equal to the design impedance. The effect of the termination value on the pole locations for the filters has also been discussed.

The great advantage of prototype image-impedance filters is the very simple design procedure involved, based on two parameters only (R_0 and ω_{c0}). The normalized 'model' filters of Fig. 2 may be readily denormalized with respect to frequency and design impedance and hence a practical filter having specified values of R_0 and ω_{c0} is very easily obtained. Providing that some degree of pass-band ripple is acceptable then it has been shown that the response of the filter may be improved by the introduction of a known amount of mismatch at the termination. The amount of mismatch is conveniently related to the pass-band ripple by the curves given (Figs. 6 and 9). Obviously, it is desirable to design for the maximum tolerance ripple in order to achieve the sharpest cut-off characteristic, but the greater the ripple the more non-linear is the phase response, see, for example, Holt.⁷ The procedure outlined in this paper is easily extended to the cases of high-pass and band-pass

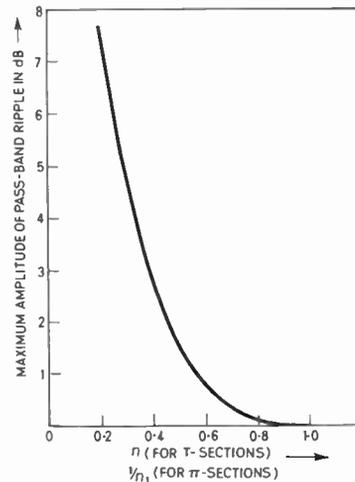


Fig. 10. Approximate value of pass-band ripple as a function of n or n_1 for two prototype sections in cascade.

filters by the well-known frequency transformation techniques.

5. Acknowledgments

The author wishes to acknowledge the interest and encouragement given by Mr. B. R. Evans, Head of the Department of Electrical Engineering, and the valuable assistance of Mr. K. Pollak with the diagrams. It is also desired to acknowledge the generous provision of computer facilities by Mr. F. J. Hawley, former Head of the Department of Mathematics and Physics, and the ready help with programming given by Mr. B. Coleman and Mr. B. J. Vokes.

6. References

1. O. J. Zobel, 'Theory and design of uniform and composite electric wave filters', *Bell Syst. Tech. J.*, 2, pp. 1-46, January 1923.
2. E. A. Guillemin, 'Communication Networks', Vols. 1 & 2. (Wiley, New York, 1935.)
3. W. Frazer, 'Telecommunications'. (Macdonald, London, 1957.)
4. J. L. Stewart, 'Circuit Theory and Design'. (Wiley, London, 1956.)
5. L. Weinberg, 'Network Analysis and Synthesis'. (McGraw-Hill, New York, 1962.)
6. S. Butterworth, 'On the theory of filter amplifiers', *Experimental Wireless and Wireless Engineer*, 7, pp. 536-41, October 1930.
7. A. G. J. Holt, 'A comparison of five methods of low-pass passive filter design', *The Radio and Electronic Engineer*, 27, pp. 167-180, March 1964.

Manuscript first received by the Institution on 22nd November 1965, in revised form on 24th May 1966, and in final form on 9th November 1966. (Paper No. 1105.)

© The Institution of Electronic and Radio Engineers, 1967

Letters to the Editor

Radar Reliability

SIR,

With reference to A. J. Harrison's recent paper 'Radar reliability on trawlers',[†] in Section 3, paragraph (b) (ii) I assume that 'receiver noise figure' should read 'system gain'. I am sure that Mr. Harrison will agree that 12 dB is hardly a small reduction in noise figure and that a receiver having a noise figure of 24 dB would be dangerously at fault. However, a reduction of 12 dB in system gain is just acceptable in terms of the specification.

The point that arises from this apparent quibble is that the point of deterioration should be located and the assessment weighted by the location. In practice such a change of performance would be due to losses spread through the system, waveguides, rotating joints, crystals, magnetrons, etc.

T. G. CLARK,
C.ENG., M.I.E.R.E.

Astaron Electronics Ltd.,
Cyldon Works, Fleets Lane,
Poole, Dorset.

31st January 1967.

SIR,

Mr. Clark's letter raises a number of interesting points. I agree that a deterioration of 12 dB in receiver noise figure must be regarded as a defect in the receiver, to be corrected as soon as possible. *In this particular case*, however, the Type Approval test measurements made by

[†] *The Radio and Electronic Engineer*, 33, No. 1, pp. 27-30, January 1967.

the Admiralty indicate that the maximum detection range of a second-class buoy would decrease from about 5 miles to 3 miles. This is roughly the same range at which such a buoy normally would be detected by one of the smaller type-approved sets produced by a number of manufacturers. It exceeds by a safe margin the figure of 2 miles, which is the minimum requirement of the Board of Trade specification. It can therefore hardly be said to be dangerous.

The choice of the receiver as the site of this fault has no particular significance. Our experience indicates that a performance loss of this order is rarely due to more than one unit. Losses due to waveguides and rotating joints do not vary more than 1 or 2 dB, except due to obvious catastrophic causes. Modern crystals fall no more than perhaps 3 dB in performance during life, while a loss of 3 dB in magnetron power would be located by the performance monitor. Changes of this order would justify replacement in the course of normal maintenance.

The wording of this paragraph was deliberately chosen to emphasize the point that a fault which leaves the radar performance reduced but still operationally useful, must be considered in a different category from a fault which renders the equipment completely useless. It would appear to have achieved its object.

A. HARRISON,
C.ENG., M.I.E.R.E.

Kelvin Hughes,
A Division of Smiths Industries Ltd.,
Selinas Lane,
Dagenham, Essex.

9th February 1967.

Correction

The heading describing the letter by Mr. F. Oakes on page 143 of the February issue of *The Radio and Electronic Engineer* contained a small but important error.

It should read 'Graphical Derivation of Trajectories for Simple R-L-C Networks containing a Non-linear Resistance Element with a Negative Resistance Region', i.e. not 'R-C Networks'.

Correspondence of a technical nature is welcomed by the Editor for consideration for publication in *The Radio and Electronic Engineer*. Writers may put forward new ideas, which perhaps are not sufficiently advanced or are too brief to form the basis of a paper or short contribution, or they may comment on papers already published, i.e. as 'written discussion'.

The Design of a Magnetic Thin-film Store for Commercial Production

By

R. S. WEBLEY, B.Sc., F. Inst.P.†

AND

A. T. GIBSON, B.Sc.‡

Summary: A magnetic thin-film store of 1024 words, each of 50 bits, has been developed which repetitively read-rewrites at 2 MHz. The 2 MHz store has operated for 5000 hours without change. A similar store of 128 words cycles at 3 MHz.

From the basic method of film operation, the element dimensions, flux linkage and signal outputs have been derived and measured. A brief description of the method of construction precedes some conclusions on the suitability of the design for commercial production.

List of Symbols

W_{\perp}	minimum permissible transverse magnetic field for writing	N	noise pulse area
D_{\perp}	maximum permissible transverse disturbing magnetic field	L	inductance/unit length of transmission line
W_{\parallel}	minimum permissible axial magnetic field	C	capacitance/unit length of transmission line
D_{\parallel}	maximum permissible axial disturbing magnetic field	Z	surge impedance of transmission line
w_a	width of address conductor	ϵ	effective permittivity of transmission line
d_a	spacing of address conductors between planes	v_0	velocity of electromagnetic waves in freespace
d_s	spacing of sense conductors between planes	C'_s	capacitance/unit length of sense line loaded with address crossing capacitances
w_s	width of sense conductor	C_s	capacitance/unit length of unloaded sense line
I_a	address current (ampere-turns)	ϵ_s	effective permittivity of the medium separating the sense conductors
δ	digit pitch	ϵ_{a-s}	effective permittivity of medium separating address from sense conductors
α	address pitch	α_{a-s}	spacing of address from sense conductors
t_m	thickness of magnetic layer	n	number of bits in a word
l	length of element	Z_d	surge impedance of digit line loaded with address crossing capacitances
l_p	distance between poles of magnetic element	C_d	capacitance/unit length of unloaded digit line
f	fraction of element flux linked with sense loop	ϵ_d	effective permittivity of medium separating address from digit conductors
B	saturation magnetic induction of film element	τ_s	sense propagation delay
L_a	inductance of address wire	τ_d	digit propagation delay
C_{a-s}	capacitance between one address conductor and one sense conductor	I_{in}, V_{in}	input and output currents and voltages of a transmission line
Z_s	surge impedance of sense-line loaded with address crossing capacitances	I_{out}, V_{out}	
τ	rise-time of leading edge of address pulse	R	resistance/unit length of a transmission line
T	amplifier response time	t	time
S	signal pulse area	V	output voltage of amplifier
		V_0	equivalent signal input voltage
		Φ	flux of magnetic element
		I	intensity of magnetization

† Research Laboratories, Electric & Musical Industries Ltd., Hayes, Middlesex.

‡ E.M.I. Electronics Ltd., Hayes, Middlesex.

1. Introduction

A word-organized magnetic film store described by Raffel¹ in 1959 used isolated circular magnetic elements deposited on thin glass bases. The suggestion was then made that improved performance and higher packing density would be obtained by the use of rectangular elements, and such a store is described in this paper. The design which has been developed for commercial production is based on printed circuit techniques, with the result that high reliability and production economies have been achieved.

For the initial design target a size of 1024 words, each of 50 bits, was chosen. It was anticipated that this would satisfy a number of applications, and would reveal the design difficulties for larger and faster stores.

2. Method of Film Operation

The remanent magnetization of films used for data storage lies in one of two opposing stable directions, representing digits '0' or '1'. To sense the magnetic state, a magnetic field transverse to the stable axis causes rotation of the magnetization within the film plane by 90°. The flux linked with a sense-loop is

reduced to zero and an output signal is generated whose polarity depends on the original direction of magnetization (Fig. 1—'read' interval). Now to record a new digit '0' or '1', the temporary addition of a small field parallel to the stable axis rotates the magnetization a little, which rotation continues upon removal of the transverse field until the magnetization again lies in the stable axis (Fig. 1—'write' interval). The transverse field is known variously as the address, word or drive field, and the axial field as the digit or information field.

During a 'write' period, elements on either side of the row of elements being written into experience:

- (i) a small field transverse to the stable axis arising from magnetic poles of elements being addressed as well as from the fringe of the word field,

and

- (ii) an information field parallel to the stable axis.

This condition of disturbance may be represented by point D in Fig. 2, in which are shown threshold orthogonal fields which cause magnetization-reversal of a typical element.² In the same figure, point W represents the writing condition.

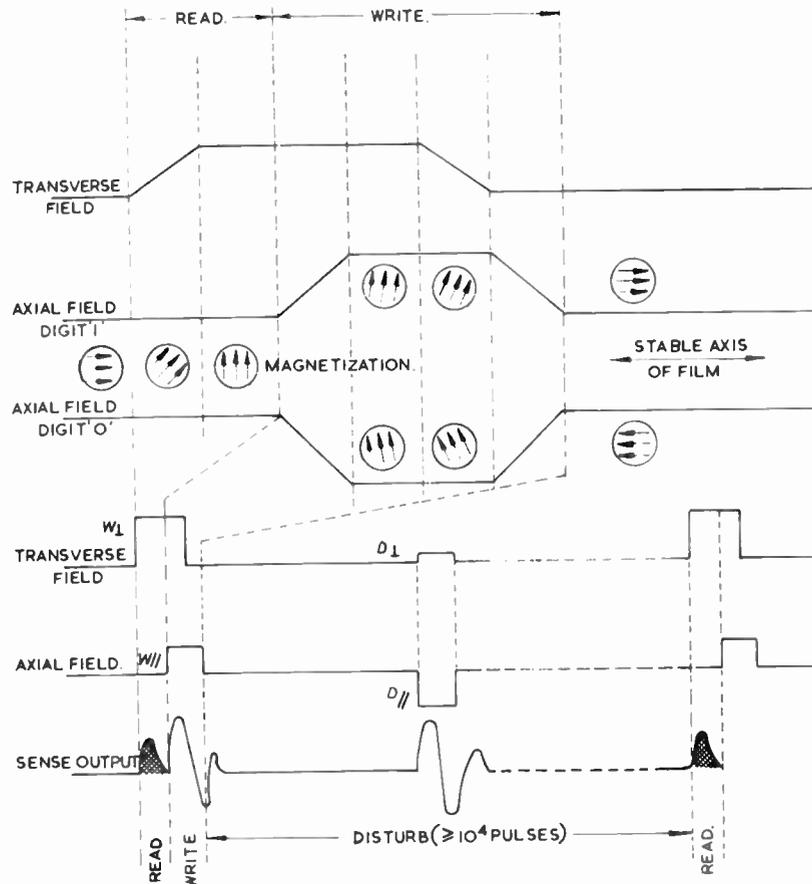


Fig. 1. Orthogonal current pulse trains used for film testing.

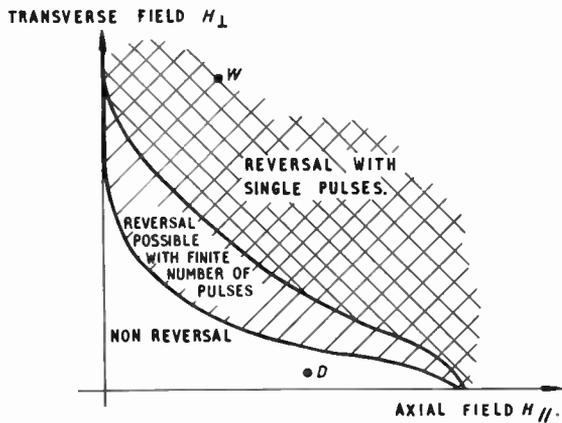


Fig. 2. Threshold curves of magnetization reversal by orthogonal fields.

For satisfactory operation, point D must be situated in the area of non-reversal, and point W should be where reversal by a single pulse occurs; this is ensured by testing individual elements with the pulse sequence shown in Fig. 1. Pulse amplitudes are given in terms of the co-ordinates of D and W, consequently W_{\perp} represents the minimum permissible transverse field, D_{\perp} the maximum permissible transverse disturbing field, and D_{\parallel} and W_{\parallel} the maximum and minimum permissible axial fields respectively. The values chosen for these co-ordinates were $W_{\perp} = 8.1$, $D_{\parallel} = 1.7$, $W_{\parallel} = 1.0$, and $D_{\perp} = 0.5$ oersteds.

3. Choice of Film and Wiring Dimensions

The wiring associated with rectangular magnetic elements is shown in Fig. 3, with symbols for the relevant dimensions. The values finally chosen permit the store to be assembled easily; its signal/noise ratio is high, the attenuation and transmission delay along digit/sense lines is kept to a minimum, and the necessary drive currents are compatible with transistor circuit requirements.

3.1. Element Width, Length and Thickness

The magnetic field midway between the address conductors is $0.8I_a/w_a \tan^{-1} w_a/d_a$ Oe, where I_a is the ampere-turns of the address current. To accommodate, without strain, the substrate supporting a magnetic film, a minimum sense conductor spacing (d_s) and hence address conductor spacing (d_a) is necessary. To generate the address field W_{\perp} using available high-speed transistors, an upper limit to w_a is then found and, for easy alignment between elements and wiring, it is undesirable to reduce w_a much below this limit. A value of 0.02 in for the element and address conductor width was therefore chosen.

The skew and dispersion of the magnetic axis are both reduced as the length l of the element is increased

(Table 1) giving an increase in digit current range ($W_{\parallel} \rightarrow D_{\parallel}$). However, too great a length decreases the signal/noise ratio. A compromise length of 0.075 in was adopted.

Table 1

	Shape	Skew	Dispersion
Circular	0.060 in diameter	+3° → -13°	4°
Rectangular	0.080 in × 0.030 in	+3° → 0°	2°
"	0.080 in × 0.020 in	-2° → 0°	1°

For a given film shape the maximum signal increases with thickness t_m , but since demagnetizing fields also increase, these require more powerful address current generators. Also as thickness increases the coercivity of the film is reduced³ leading to a smaller digit current range. In practice, yields of at least 80% are normally obtained where the pulse sequence of Fig. 1 is used to test plates of 80 elements, and the threshold

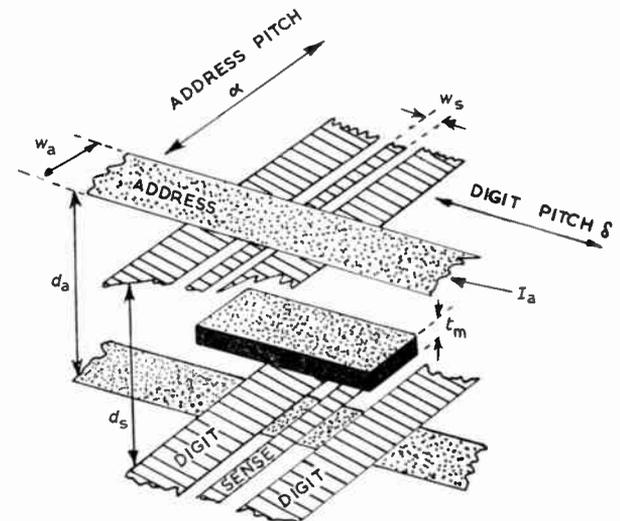


Fig. 3. Local arrangement of film element and wiring.

criterion is signal amplitude = 3×10^{-11} volt. second on a simple pickup loop. Film thickness may vary from 1200 to 1400 Å.

3.2. Wiring Spacings

To support the magnetic elements, microscope cover slides of nominal thickness 0.005 in are used. Since most of these are warped, it was found necessary to allow 0.009 in spacing if magnetostriction effects, even with 82% Ni -18% Fe, were to be avoided. The digit/sense wiring is placed close to the glass plates, to reduce digit/sense coupling: early

experiments had shown the desirability of this, since recovery time of the sense system after a digit pulse was applied occupied a major part of the complete cycle for a large store.

The spacing d_a of the address wires needs to be as small as possible to minimize the field acting on neighbouring elements. However, the unwanted voltage transient arising from address to sense capacitance is then large, causing a deterioration of the signal/noise ratio. The compromise address line spacing chosen was 0.018 in, requiring a drive of 750 mA-turns which is generated from a 2-turn address loop carrying a current of 375 mA.

3.3. Pitch of Digit and Address Wires

The smallest address pitch is determined by the worst case magnetic field of surrounding elements, and the fringe field of adjacent address lines. The total transverse disturbing field as shown in Appendix 1, is given by

$$H_{\perp} \approx \frac{0.07 w_a + 0.2 I_a d_a}{\alpha^2} \text{ Oe} \quad \dots\dots(1)$$

where α is the address pitch. For the chosen values $\alpha = 0.073$ in, $w_a = d_a = 0.02$ in, and $I_a = 0.75$ A-turns, we find $H_{\perp} = 0.33$ Oe. The transverse field D_{\perp} (Figs. 1 and 2) tested for is 0.5 Oe, providing an adequate safety margin in practice.

The digit pitch $\delta = 0.09$ in is determined by the length of the elements, 0.075 in, with a separation of 0.015 in.

The resultant worst-case axial disturbing field applied to a non-selected element, due to all the surrounding elements as shown in Appendix 2 is given by

$$H_{\parallel} \approx 0.085 w_a \left(\frac{1}{\alpha^2} + \frac{1}{\delta^2} \right) = 0.21 \text{ Oe} \quad \dots\dots(2)$$

For a selected element this reduces to

$$H_{\parallel} = \frac{0.085 w_a}{\alpha^2} = 0.125 \text{ Oe} \quad \dots\dots(2a)$$

since there is no contribution to the axial field from neighbouring elements in the same address row. The effect of these fields is to reduce the digit range of a practical store to 110 → 150 mA.

3.4. The Digit/Sense Wiring

The arrangement employed is shown in Fig. 3 in which a pair of co-planar half-digit lines is printed adjacent to the sense line. To minimize interaction between sense and digit, a change of phase is then made every 64 words. Also, to use the contra-directional coupling of coupled transmission lines,⁴ the digit current is introduced at the end of the store opposite to the sense amplifier, and both lines are terminated

in their characteristic impedance at the free ends. Further reduction of the digit to sense noise, and hence amplifier recovery time, can be made by localized alteration of the mutual coupling.

4. Store Parameters

Having chosen the structural dimensions, accurate mathematical expressions can be obtained for the electrical properties, and the parameters affecting the complete memory system design can be ascertained.

4.1. Efficiency of Flux Linkage with Sense Wire and the Signal/Noise Ratio

It can be shown that the fraction of element flux linked with the sense loop is $f = 2/\pi \tan^{-1} l_p/d_s$, where l_p is the distance between the poles of the magnetic element. In Fig. 4 the signal from an element of length 0.075 in, is plotted against the centre sense spacing d_s , and superimposed is a theoretical curve for $l_p = 0.045$ in. The linkage efficiency, for the dimensions chosen, is approximately 80%.

The signal pulse area $S = f w_a t_m B \cdot 10^{-8}$ Vs. If the pulse front of address current is approximated by a ramp function (Fig. 5), and the address lines are represented by an inductance, the capacitively coupled sense noise voltage may be represented by two bipolar spikes, each having an area $L_a C_{a-s} Z_s I_a / 2\tau$, which are then integrated by the amplifier. For $\tau > T$ (the amplifier response time) the signal/noise ratio is

$$S/N = \frac{2 f w_a t_m B \tau \cdot 10^{-8}}{L_a C_{a-s} I_a Z_s} \quad \dots\dots(3)$$

The parameters of balanced and open strip lines (Figs. 6(a) and 6(b)) have been rigorously determined for lines having zero thickness.⁵ The smallest value

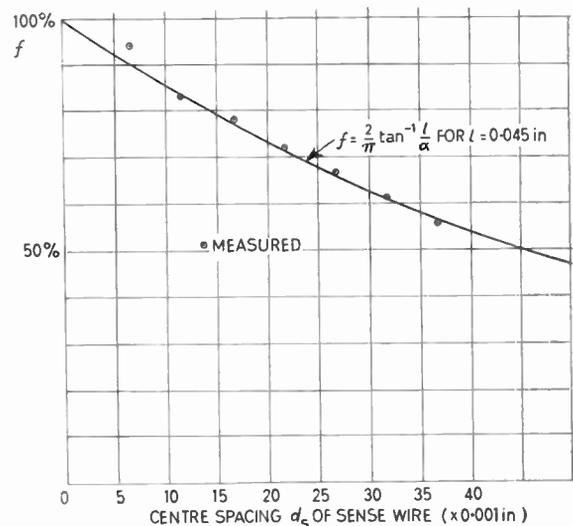


Fig. 4. Signal amplitude as a function of sense-wire spacing.

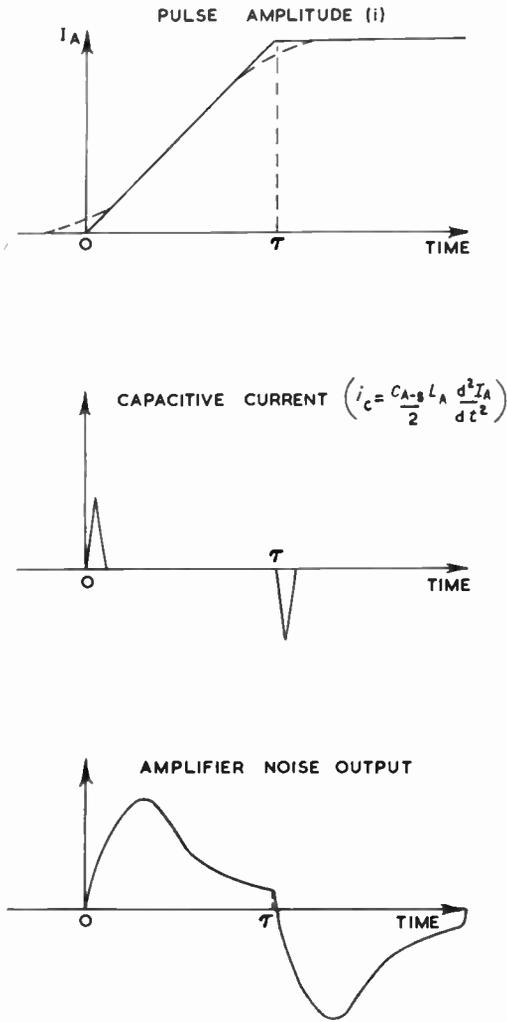


Fig. 5. Leading-edge of address-current pulse and derived noise.

of width/thickness ratio used is 5 : 1, consequently the error involved by assuming zero thickness is considered to be negligible. For strip lines having width/separation (w/d) within the range $1 \leq w/d \leq 8$, the capacitance per unit length can be accurately represented by the following expressions:

balanced line

$$C = \epsilon \{ (0.096 w/d) + 0.088 \} \text{ pF/cm} \quad \dots\dots(4)$$

open line

$$C = \epsilon \{ (0.11 w/d) + 0.14 \} \text{ pF/cm} \quad \dots\dots(5)$$

Using the appropriate relationship for C , the inductance and impedance of a line is given by⁶

$$L = \frac{\sqrt{\epsilon}}{v_0^2 C} \text{ H/cm} \quad \dots\dots(6)$$

$$Z = \frac{\sqrt{\epsilon}}{v_0 C} \text{ ohm} \quad \dots\dots(7)$$

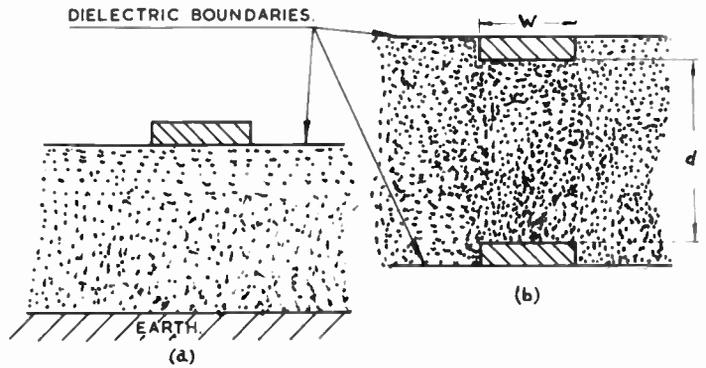


Fig. 6. Strip line configurations. (a) Open strip. (b) Balanced strip.

where v_0 is the velocity of electromagnetic waves in free space, and ϵ the effective permittivity of the medium. The cross-coupling capacitors C_{a-s} , modify the sense line impedance thus:

$$Z_s = \sqrt{\frac{L_s}{C'_s}} \text{ ohm} \quad \dots\dots(8)$$

where

$$C'_s = C_s + \frac{C_{a-s}}{2\alpha} \text{ pF/cm} \quad \dots\dots(9)$$

and α is the word line pitch.

Combining these equations

$$Z_s = \frac{1}{v_0 C_s} \left\{ \frac{\epsilon_s}{1 + \frac{C_{a-s}}{2\alpha C_s}} \right\}^{\frac{1}{2}} \text{ ohm} \quad \dots\dots(10)$$

The value of C_{a-s} can be calculated using eqn. (5) and is equal to the sum of a length w_a of an open strip width w_s and a length w_s of an open strip width w_a , less the cross-coupling capacitance neglecting fringe effects.

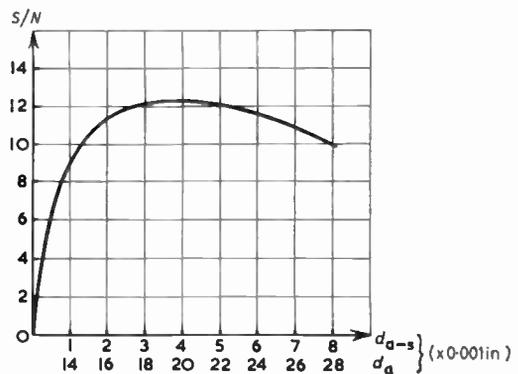


Fig. 7. Theoretical signal/noise of store as a function of address-wire spacing.

$$C_{a-s} = \epsilon_{a-s} \left[0.132 \frac{w_a w_s}{d_{a-s}} + 0.14 (w_a + w_s) \right] \text{ pF}$$

For the materials and dimensions chosen, a 50 bit word length, and $\tau = 40 \text{ ns}$, a signal/noise ratio of 12 : 1 results.

In Fig. 7 this ratio is plotted against the spacing between address and sense conductors d_{a-s} .

4.2. Line Impedance and Propagation Delay

The line impedances can be calculated using eqns. (4) to (10).

Address line inductance L_a

$$= \frac{4 n \delta 10^{12}}{v_0^2 \{ (0.096 w_a / d_a) + 0.088 \}} \text{ H} \quad \dots\dots(11)$$

Sense line impedance Z_s is given by eqn. (10).

Digit line impedance Z_d

$$= \frac{1}{v_0 C_d} \left\{ \frac{\epsilon_d}{1 + (C_{a-d} / 2\alpha C_d)} \right\}^{\frac{1}{2}} \text{ ohm} \quad \dots\dots(12)$$

For the dimension chosen, and $n = 50$, $L_a = 0.27 \text{ } \mu\text{H}$, $Z_s = 80 \text{ ohms}$ and $Z_d = 36 \text{ ohms}$.

The sense line propagation delay

$$\tau_s = \sqrt{L_s C_s} \quad \dots\dots(13)$$

Combining this with eqns. (8), (9) and (10) gives

$$\tau_s = \frac{\alpha}{v_0} \left\{ \epsilon_s \left(1 + \frac{C_{a-s}}{2\alpha C_s} \right) \right\}^{\frac{1}{2}} \text{ s/word} \quad \dots\dots(14)$$

Similarly the digit line propagation delay is

$$\tau_d = \frac{\alpha}{v_0} \left\{ \epsilon_d \left(1 + \frac{C_{a-d}}{2\alpha C_d} \right) \right\}^{\frac{1}{2}} \text{ s/word} \quad \dots\dots(15)$$

The difference between these two delays is negligible and substituting values produces $\tau_s = \tau_d = 5.5 \text{ ps/word}$.

For a 1024 word store, the propagation delay is calculated to be 16 ns. However, the store is constructed using planes containing 64 words, so there is an additional delay in the plane to plane interconnection, which amounts to a further 14 ns.

4.3. Line Attenuation

The attenuation of a signal pulse propagating along a transmission line is given by $\exp(-R/2Z)$, where Z is the characteristic impedance of the line. For current rise times at present in use, skin effect distortion is negligible, hence R is the low frequency resistance of the line.

The attenuation calculated for the digit and sense lines of a 1024 word store is 5% and 10% respectively, which is in good agreement with measured results.

5. Store Interrogation

For a word-organized store, one of the requirements is a cheap form of address current selection with individual circuit elements mounted as part of the memory structure, to reduce the number of interface connections. The design of the interrogation circuits is now discussed.

5.1. Address Selection

The system used is shown in Fig. 8, and this employs 1 : 1 transformers for isolation with a diode in the secondary so that the address current is substantially independent of the repetition rate.

Small ferrite beads are used for the transformer core (3 mm long and 3 mm external diameter) with winding ratio 3 : 3. The drive current insertions loss is given by $L_a / (L_p + L_a)$ where L_p is the primary inductance, and in the worst case the loss is equal to 10%. Also the drive voltage requirements are increased by a factor of 2 due to the leakage inductance.

5.2. Sensing

To take advantage of the bi-polarity of output signal, it is sampled at the time of its peak value. Two tunnel diodes form a very efficient strobe circuit^{7, 8} and using 2 mA tunnel diodes, and allowing for

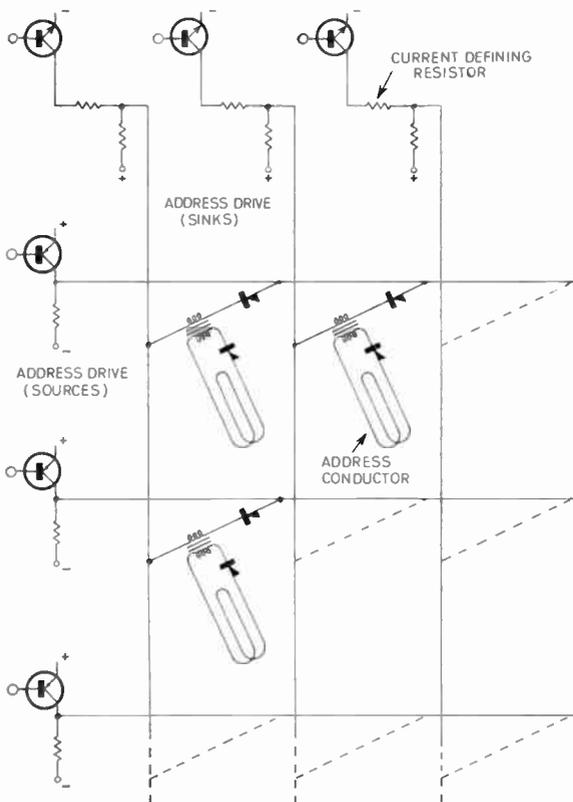


Fig. 8. Address selected system.

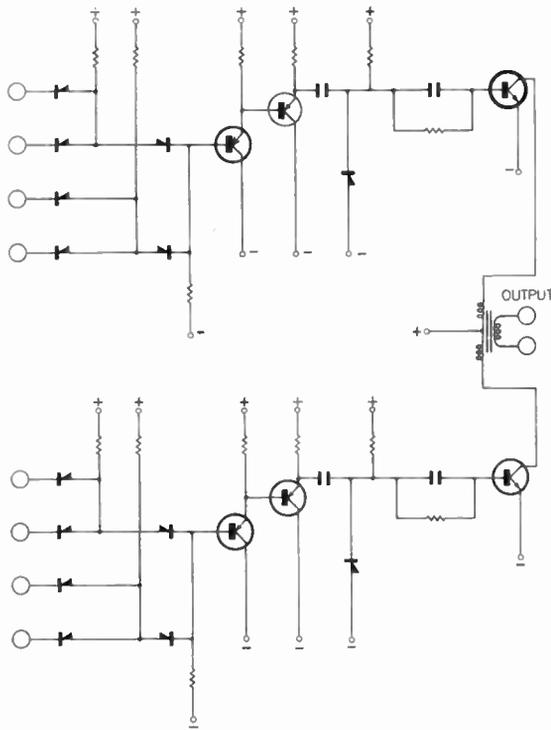


Fig. 9. Digit drive circuit.

± 3% rail variations and a temperature range 20°C to 50°C, the sensitivity is better than 30 μA.

In the 1024 word store the time uncertainty between signal and strobe delays is 45 ns, and for effective strobing the amplified signal has to be wide enough to cater for this time uncertainty.

The amplified signal approximates to

$$V = V_0 \frac{t}{T} \exp\left(1 - \frac{t}{T}\right).$$

The input signal is a relatively narrow spike, but the equivalent amplitude V_0 of a signal of the same shape as the output may be calculated:

$$\Phi f = \int_0^{\infty} V_0 \frac{t}{T} \exp\left(1 - \frac{t}{T}\right) dt = eV_0 T$$

giving

$$V_0 = \frac{\Phi f}{eT}$$

where Φ is flux of the element.

In the practical circuit $T = 16$ ns, giving a peak signal amplitude of 600 μV for the thinner layers of 1200 Å, but to allow for signal variation, the amplifier has been designed to deliver ± 60 μA into the strobe circuit for a signal strength of ± 250 μV. The circuit incorporates a differential input stage for common mode noise reduction.

5.3. Digit Drive

The nominal digit current pulse of 130 mA, must be of a polarity depending on the state of the corresponding information register stage, and a suitable circuit is shown in Fig. 9. An important requirement is that the absolute current during the write period is substantially independent of repetition rate. This condition is fulfilled but if the mark-to-space ratio is equal to unity, the output transistors must be capable of providing twice the nominal current, i.e. 260 mA.

6. Construction of Store

Film elements are placed between two-layer printed boards, and the arrangement is stabilized by impregnating with petroleum jelly. The board pairs are stacked and interconnected in the digit/sense direction and to the address matrix with printed connectors. In all cases relatively long lengths of matrix conductors are plated with tin-lead solder, and groups of conductors are united in pairs using a heated block. Some 30,000 joints of this type are needed in the 1024 word store.

Address matrix components are mounted on printed boards for easy servicing, and pulse current is carried to these via transmission lines to facilitate high speed working. A slotted copper screen separates the matrix and storage planes and interconnecting leads pass through the slots.

A mu-metal box (11 in × 10 in × 8 in) shields the elements from external magnetic fields.

7. Store System and Performance

A block diagram of the store system, which uses diode-transistor logic, is shown in Fig. 10. A typical propagation delay through a logic unit is 25 ns, and rise times are generally faster than 15 ns. The measured access time of the 1024 word/50 bit store was 270 ns, and the read-rewrite cycle time was 500 ns, these times being made up as follows:

(a) time to decode address	90	access time	270 ns
(b) address drive circuit delay	20		
(c) address current rise-time	40		
(d) signal amplifier delay	30		
(e) sense propagation time	30		
(f) strobe delay	20		
(g) word register delay + rise-time	40		
(h) digit circuit delay	20	cycle time	500 ns
(i) digit current width	120		
(j) amplifier recovery time from digit noise	200		

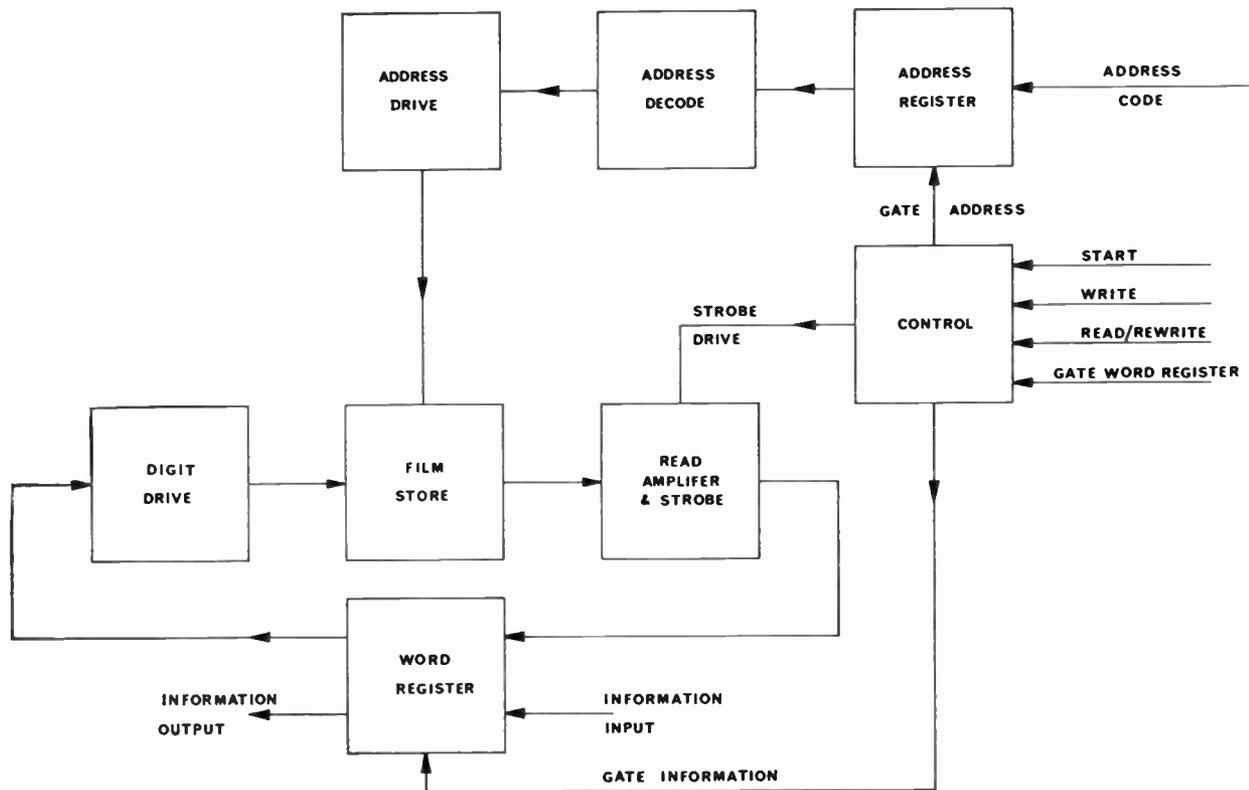


Fig. 10. Block diagram of store system.

A typical amplifier output is shown in Fig. 11 and the recovery (j) is clearly seen. Figure 12(a) illustrates the safety margin of the store system to combat drift, and in Fig. 12(b) the strobe threshold is plotted against digit current showing the effect of a large number of disturb cycles.

A smaller size of store should be faster, and the first production system having 128 words of 16 bits has an access time of 220 ns, with read-rewrite cycle time 330 ns. The amplifier output, while all words are cycled consecutively, is seen in Fig. 13. The apparent indecision immediately following the signal arises from the use of two short address pulses in place of one relatively long one to reduce heating of transistors. However, the availability of an address drive transistor with less storage would enable a cycle time of less than 300 ns to be achieved.

8. Conclusions

A magnetic thin-film store, having a worthwhile advance in speed over currently available stores has been successfully demonstrated. Higher operational speed will be achieved as semiconductor techniques improve. Batch production and testing of the magnetic elements, together with the use of well-established printed circuit techniques and a simple reliable

method of interconnections, forms the basis of a cheap commercial product.

Experience gained in the small scale production has shown that by simple design modification, larger stores can be made having cycle times in the region of 250 ns, at a cost comparable with a 1 μ s core store.

9. Acknowledgments

It is a pleasure to acknowledge the contributions of Messrs. A. C. Moore, A. S. Young and R. J. Heritage who have collaborated with us freely for many years, and who have made many measurements on our behalf.

We also thank the Directors of Electric & Musical Industries, Ltd. for permission to publish this paper.

10. References

1. J. Raffel, UNESCO Conference, Paris, April 1959. (Paper No. K6).
2. W. Dietrich and W. E. Proebster, 'A study of switching in thin magnetic films', International Solid State Circuits Conference, 1961.
3. M. Prutton, 'Thin Magnetic Films', p. 96. (Butterworth, London, 1964).
4. B. M. Oliver, 'Directional electromagnetic couplers', *Proc. Inst. Radio Engrs*, 42, No. 11, pp. 1186-92, November 1954.

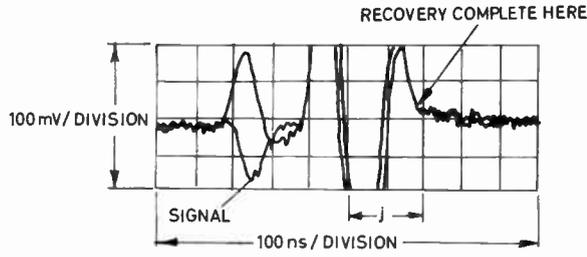


Fig. 11. Typical amplified signal.

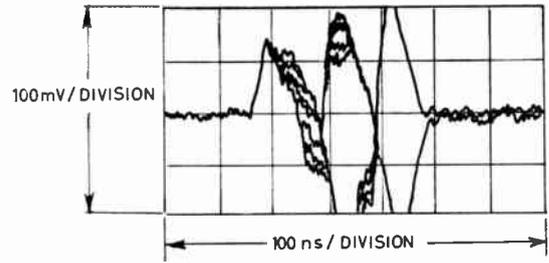


Fig. 13. Amplifier output while all words are cycled consecutively.

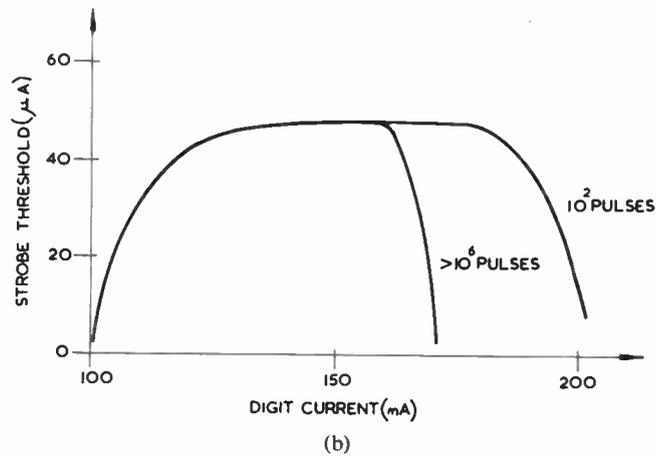
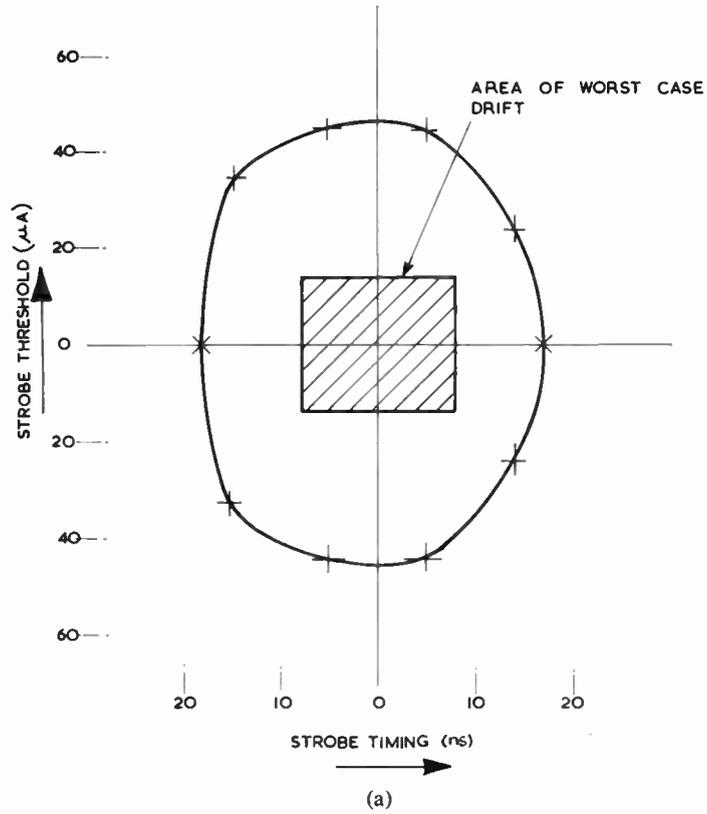


Fig. 12. Safety margin of the store to drift.

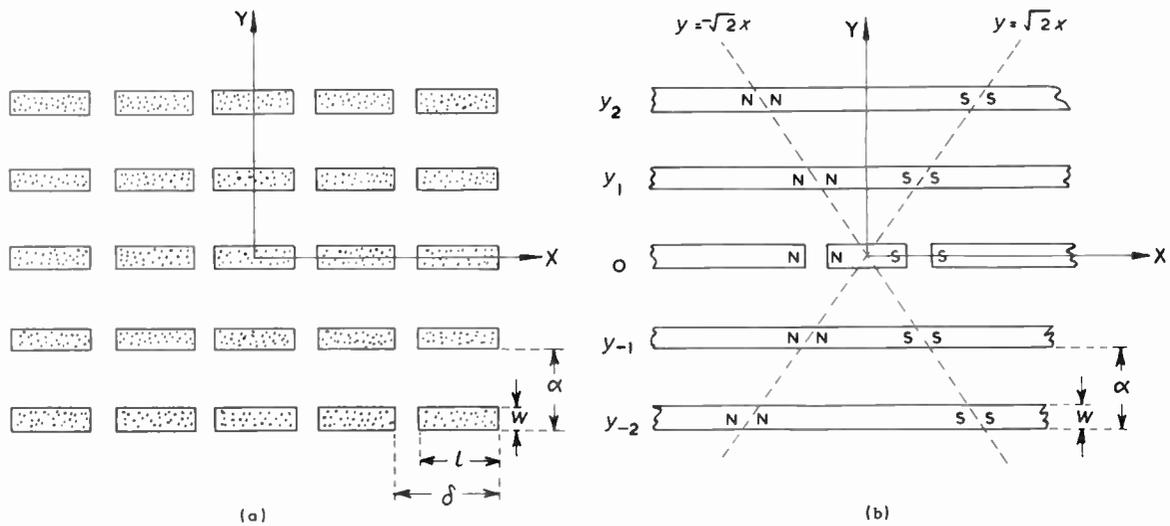


Fig. 14. (a) Array of film elements. (b) Simulated array.

5. K. G. Black and T. Higgins, 'Rigorous determination of the parameters of microstrip lines', *Trans. I.R.E. on Microwave Theory and Technique*, MTT-3, p. 93, March 1955.
6. M. Arditi, 'Experimental determination of the properties of microstrip components', *Elect. Commun.*, 30, p. 283, 1953, and *Nat. Conv. Rec. I.R.E.*, Pt. 10, p. 27, 1953.
7. E. Goto *et al.*, 'Esaki diode high-speed logical circuits', *Trans. I.R.E. on Electronic Computers*, EC-9, No. 1, pp. 25-29, March 1960.
8. G. H. Perry and E. W. Shallow, 'A read-out circuit for high-feed non-destructively read stores', *Proc. I.F.I.P. Congress*, p. 597, 1962.

11. Appendix 1

Worst-case Hard-axis Field

The simplifying assumption is made that the discrete films of Fig. 14(a) are replaced with horizontal strips of the same width and thickness, see Fig. 14(b).

The worst transverse disturbing field at the origin occurs when (i) strips y_2, y_3, y_4 , etc., are magnetized with N-poles on the Y axis, (ii) strips y_{-1}, y_{-2}, y_{-3} , etc., are magnetized with S-poles on the Y axis, and (iii) the magnetism of strip y_1 , is continually swung between the X and Y axes. The field of (i) and (ii) is static and is given by:

$$\frac{4Itw}{\alpha^2} \left[\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots \right] - \frac{2Itw}{\alpha^2}$$

$$\approx \frac{4Itw}{\alpha^2} \left[\frac{\pi^2}{6} - \frac{1}{2} \right] = 0.05 \frac{w}{\alpha^2} \quad \text{for } I = 900$$

$t = 1200 \text{ \AA}$

The dynamic field of (iii) arises from that of the film

and the fringe field of the conductors, and has the value

$$\frac{0.022 w}{\alpha^2} + \frac{0.2 I_a d_a}{[\alpha - (w/2)]^2}$$

(units of ampere and cm).

12. Appendix 2

Worst-case Easy-axis Field

Again using the premise of Appendix 1, the easy-axis field at (0,0) is greatest when poles are situated on the lines $y = \pm \sqrt{2}.x$. Its value for all strips except y_0 is:

$$\frac{8It}{\sqrt{3}} \left[\frac{w}{\alpha^2 - (w/2)^2} + \frac{w}{(2\alpha)^2 - (w/2)^2} + \frac{w}{(3\alpha)^2 - (w/2)^2} + \dots \right]$$

$$= 0.084 \frac{w}{\alpha^2} \quad \text{for } I = 900$$

$t = 1200 \text{ \AA}$

$\alpha^2 \gg (w/2)^2$

The field of the nearest neighbours in y_0 is $8Itw/\delta^2$, so that the total static field is:

$$0.085 w \left(\frac{1}{\alpha^2} + \frac{1}{\delta^2} \right) = 0.21 \text{ Oe}$$

for $w = 0.02$ in

$\alpha = 0.073$ in

$\delta = 0.09$ in

Manuscript first received by the Institution on 19th August 1965 and in revised form on 14th July 1966 (Paper No. 1106/C94)

© The Institution of Electronic and Radio Engineers, 1967

Synthesis of a Cascaded Network by Transfer Matrix Factorization

By

M. S. HUNT, M.Sc.†

Summary: This paper discusses the use of transfer matrices in the synthesis of cascaded electronic networks. Previous work on this subject has been performed by trial and error methods, but here it is shown that for certain special classes of transfer matrix, the solution, if indeed a solution exists, can be predetermined logically.

1. Introduction

There are two basic problems in electronic network theory: circuit analysis, in which the response of a given network is to be found; and circuit synthesis, in which the circuit with a given response is to be found. Most methods of synthesis are based on relating voltage with current by means of the so-called z -matrix, or impedance matrix; the method used below, however, is based on relating input with output by means of the transfer matrix, or A -matrix.

The transfer-matrix method of synthesis of a network may be expressed in two steps, as:

- (i) to derive a transfer matrix corresponding to a specified response;
- (ii) to factorize this matrix into a combination of simpler matrices, each of which represents an element of a cascaded network.

Step (i) is well defined and further discussion of it is omitted here. The methods put forward by Davies^{1, 3} for step (ii) tend to be tentative and rather vague, but it will be shown that they can be made definite in many cases.

2. The Transfer Matrix

If V_i , I_i are the input voltage and current, and V_o , I_o the output voltage and current of a network, then the transfer matrix, A , of that network is defined by

$$\begin{bmatrix} V_i \\ I_i \end{bmatrix} = A \begin{bmatrix} V_o \\ I_o \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} V_o \\ I_o \end{bmatrix}$$

For example, the transfer matrix of an impedance Z is

$$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$$

and of an admittance Y :

$$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$$

In the rest of this paper, only passive, reciprocal networks consisting of resistors, capacitors, inductors

† Computation Laboratory, University of Southampton.

and transformers will be considered. For such networks, the determinant of the transfer matrix is unity²; further, for each single component the transfer matrix takes one of the following basic forms:

$$A_z = \begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$$

for an impedance Z ;

$$A_y = \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$$

for an admittance Y ;

$$A_T = \begin{bmatrix} n & 0 \\ 0 & 1/n \end{bmatrix}$$

for a transformer of turns ratio n .

These three matrices will in future be referred to as elementary matrices.

It is a well-known property of the A -matrix that the transfer matrix of a cascaded network is the product of the transfer matrices of the elements of the cascade, in the same order.^{4, 5} Hence a cascaded, passive, reciprocal network will have a transfer matrix which may be expressed as a product of elementary matrices; for example

$$A = A_{z1} \times A_{z2} \times A_{y1} \times A_{z3} \times A_{y2} \times A_T$$

This example may represent the network shown in Fig. 1.

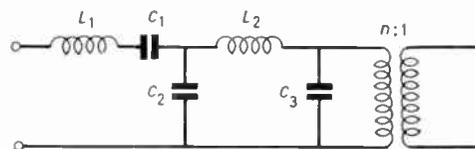


Fig. 1. Example of a cascaded passive reciprocal network.

It is easily seen that since all immittances (i.e. impedances or admittances) are functions of frequency, the transfer matrix will be a function of frequency.

To avoid possible confusion arising from using the complex quantity $j\omega$, immittances will be written in terms of $s = j\omega$. Then the above A -matrix may be written as

$$A = \begin{bmatrix} 1 & L_1s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/(C_1s) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ C_2s & 1 \end{bmatrix} \times \\ \times \begin{bmatrix} 1 & L_2s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ C_3s & 1 \end{bmatrix} \begin{bmatrix} n & 0 \\ 0 & 1/n \end{bmatrix}$$

To make the multiplication easier, introduce some normalized values for the components (for a discussion of the normalization of component values for this type of problem the reader is referred to Ref. 3) e.g. $L_1 = 1$, $C_1 = 0.5$, $C_2 = 1$, $L_2 = 2$, $C_3 = 3$, $n = 4$ to give

$$A = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2s^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \begin{bmatrix} 1 & 2s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3s & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \\ = \begin{bmatrix} 1 & 2s^{-1} + s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2s \\ s & 2s^2 + 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 12s & \frac{1}{4} \end{bmatrix} \\ = \begin{bmatrix} 3 + s^2 & 2s^{-1} + 7s + 2s^3 \\ s & 1 + 2s^2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 12s & \frac{1}{4} \end{bmatrix} \\ = \begin{bmatrix} 36 + 88s^2 + 24s^4 & \frac{1}{2}s^{-1} + \frac{7}{4}s + \frac{1}{2}s^3 \\ 16s + 24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$$

Now all elements of this matrix are rational functions of s . This is true for any transfer matrix which is created by multiplying several elementary matrices together. Hence, the general transfer matrix can be written in the form

$$A = \frac{1}{g(s)} \begin{bmatrix} B_{11}(s) & B_{12}(s) \\ B_{21}(s) & B_{22}(s) \end{bmatrix}$$

where $g(s)$, $B_{11}(s)$, $B_{12}(s)$, $B_{21}(s)$ and $B_{22}(s)$ are all polynomials in s . (This form of the A -matrix will be used in classifying problems.) For example, the previous A -matrix may be written as

$$A = \frac{1}{4s^2} \begin{bmatrix} 144s^2 + 352s^4 + 96s^6 & 2s + 7s^3 + 2s^5 \\ 64s^3 + 96s^5 & 2s^2 + 2s^4 \end{bmatrix}$$

It can be shown that the determinant of this matrix is unity.

At this stage note the following points which are true for transfer matrices of all cascaded networks which are constructed from passive components:

- (i) The degrees of polynomial elements in the same row, or column, differ by at most one; in the above example it is always one, but for networks containing resistors, they may be of the same degree.
- (ii) When the degree of $B_{11}(s)$ is greater than the degree of $B_{21}(s)$, the degree of $B_{12}(s)$ is also greater than the degree of $B_{22}(s)$.

- (iii) When the degree of $B_{11}(s)$ is less than the degree of $B_{21}(s)$, the degree of $B_{12}(s)$ is also less than the degree of $B_{22}(s)$. (Obviously, this cannot be shown by the above example since it already shows (ii).)
- (iv) For the transfer matrix of a lossless network (no resistance), $B_{11}(s)$ and $B_{22}(s)$ are even functions of s and $B_{21}(s)$ and $B_{12}(s)$ are odd functions of s .

3. The Method of Factorization

It has already been shown that a transfer matrix of a cascaded network may be written as a product of elementary matrices. Thus, for example:

$$A = A_{z_1} \times A_{z_2} \times A_{y_1} \times A_{z_3} \times A_{y_2} \times A_T$$

and the problem on hand is to find these elementary matrices given the overall transfer matrix. Now

$$A_z^{-1} = \begin{bmatrix} 1 & -z \\ 0 & 1 \end{bmatrix}, A_y^{-1} = \begin{bmatrix} 1 & 0 \\ -y & 1 \end{bmatrix} \text{ and } A_T^{-1} = \begin{bmatrix} 1/n & 0 \\ 0 & n \end{bmatrix}$$

(these may easily be verified by calculating $A_z \times A_z^{-1}$ etc.), so premultiplication of A , in the above expansion, by A_z^{-1} gives

$$A_z^{-1} \times A = A_z^{-1} \times A_{z_1} \times A_{z_2} \times A_{y_1} \times A_{z_3} \times A_{y_2} \times A_T$$

If now z is chosen to be equal to z_1 , i.e. that $A_z^{-1} = A_{z_1}^{-1}$, then

$$A_1 = A_{z_1}^{-1} A = A_{z_2} \times A_{y_1} \times A_{z_3} \times A_{y_2} \times A_T$$

This process is repeated for A_1 to give a reduced matrix A_2 , and so on until the reduced matrix is itself an elementary matrix. The synthesized network is then given by the elementary matrices whose inverses were used in the reduction process, i.e. the above example would give:

first step

$$A_{z_1}^{-1} A \text{ giving first component } z_1$$

second step

$$A_{z_2}^{-1} A_{z_1}^{-1} A \text{ giving second component } z_2$$

third step

$$A_{y_1}^{-1} A_{z_2}^{-1} A_{z_1}^{-1} A \text{ giving third component } y_1$$

and so on; the network may be drawn as shown in Fig. 2.

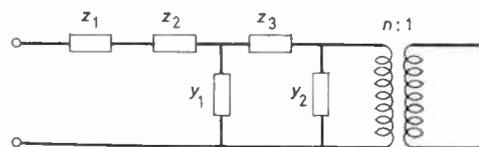


Fig. 2. Network derived in demonstrating factorization procedure.

The first part of the problem is therefore to determine the type of elementary matrix (A_z , A_y or A_T) to use in the factorization at any stage. Once this is decided upon, it is then necessary to find the type of component, and the value of the component, to be represented by this elementary matrix. For example, z_1 may be a series inductance, of impedance sL , or a series capacitance, of impedance $1/sC$.

4. Simple Lossless Networks

A simple lossless network is here defined as one whose transfer matrix is such that $g(s) \equiv 1$ when that matrix is written in the general form defined previously. For example,

$$A = \begin{bmatrix} 4+40s^2+24s^4 & \frac{3}{4}s + \frac{1}{2}s^3 \\ 16s+24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$$

The problem is now to reduce this given matrix to a product of elementary matrices. First, it must be decided which of the elementary matrices is to be used in the first stage of the factorization; this problem is eased slightly by the fact that if a transformer is included in the network, it may be assumed to appear as the last element of the network since movement of a transformer in such a network only affects the values of the components and not their type. Hence only the inverse elementary matrices A_z^{-1} and A_y^{-1} need to be considered in the factorization process. This means either

(i) $\begin{bmatrix} 1 & -z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4+40s^2+24s^4 & \frac{3}{4}s + \frac{1}{2}s^3 \\ 16s+24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$

or

(ii) $\begin{bmatrix} 1 & 0 \\ -y & 1 \end{bmatrix} \begin{bmatrix} 4+40s^2+24s^4 & \frac{3}{4}s + \frac{1}{2}s^3 \\ 16s+24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$

where z may be either Ls or $1/Cs$ and y may be either Cs or $1/Ls$.

In Section 3, it was shown that the factorization is performed in such a way that the first elementary matrix of the given matrix is completely nullified by the premultiplication process. This is ensured if the degree of one element of the given matrix is reduced by the process, in which case the degree of the other element in the same row is also reduced.¹

Hence, using (ii) above would mean reducing the degree of $A_{21}(s)$, since $A_{11}(s)$ and $A_{12}(s)$ are unaffected by this premultiplication. As $A_{21}(s)$ is already of lower degree than $A_{11}(s)$ by a factor of one, the reduced $A_{21}(s)$ would differ in degree by more than one from the corresponding $A_{11}(s)$. This means that further reduction would be impossible since it would be impossible to select a factor to cancel the appropriate term of $A_{11}(s)$.

Thus, assuming that a solution exists, only (i) may be used in this case, and the problem reduces to finding the form of the unknown z . Now if $z = 1/Cs$,

the premultiplication would yield, in the above example, a new $A_{11}(s)$ defined as

$$A_{11}(s) = (4 + 40s^2 + 24s^4) - \frac{1}{C} s^{-1}(16s + 24s^3)$$

and it is then impossible to nullify the term in s^4 . This means that an elementary matrix is not completely removed from the given matrix and the factorization is not successful. This leaves only $z = Ls$, and this gives

$$\begin{bmatrix} 1 & -Ls \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4+40s^2+24s^4 & \frac{3}{4}s + \frac{1}{2}s^3 \\ 16s+24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix} \\ = \begin{bmatrix} (4+40s^2+24s^4) - Ls(16s+24s^3) & \\ & \end{bmatrix} \\ \begin{bmatrix} & (\frac{3}{4}s + \frac{1}{2}s^3) - Ls(\frac{1}{4} + \frac{1}{2}s^2) \\ & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$$

and a choice of $L = 1$ reduces the degrees of $A_{11}(s)$ and $A_{12}(s)$ to give

$$\begin{bmatrix} 4+24s^2 & \frac{1}{2}s \\ 16s+24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$$

The arguments used above may also be applied to any simple lossless network problem, and may be summarized as follows:

- (1) If the degree of $A_{11}(s)$ is greater than the degree of $A_{21}(s)$, then the factorization must be performed using an inverse impedance matrix A_z^{-1} where the element z represents a series inductance of impedance Ls .

By similar arguments to those used above, it can be shown that:

- (2) If the degree of $A_{11}(s)$ is less than the degree of $A_{21}(s)$, then the factorization must be performed using an inverse admittance matrix A_y^{-1} where the element y represents a shunt capacitance of admittance Cs .

The problem above may now be solved to give:

$$\begin{bmatrix} 1 & 0 \\ -Cs & 1 \end{bmatrix} \begin{bmatrix} 4+24s^2 & \frac{1}{2}s \\ 16s+24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix}$$

Choosing $C = 1$, we get:

$$\begin{bmatrix} 4+24s^2 & \frac{1}{2}s \\ 12s & \frac{1}{4} \end{bmatrix}$$

Now the degree of $A_{11}(s)$ is greater than that of $A_{21}(s)$ so the next factor is A_z^{-1} ; choice of $z = Ls = 2s$ gives

$$\begin{bmatrix} 4 & 0 \\ 12s & \frac{1}{4} \end{bmatrix}$$

which in turn may be broken down into

$$\begin{bmatrix} 1 & 0 \\ 3s & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

The factorized matrix is therefore:

$$\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \begin{bmatrix} 1 & 2s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3s & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

and the network is as shown in Fig. 3.

5. Simple Lossy Networks

In the case of lossy networks, all elements of the transfer-matrix are general polynomials in s ; for example:¹

$$A = \begin{bmatrix} 1+2s+4s^2+6s^3 & 1+6s+10s^2+6s^3 \\ 2+s+6s^2 & 3+7s+6s^2 \end{bmatrix}$$

This type of problem is similar to the previous type inasmuch that if the degrees of $A_{11}(s)$ and $A_{21}(s)$ are different, the above ruling is still true; for example, the above problem must first be premultiplied by A_z^{-1} with $z = Ls$; in fact $L = 1$ gives

$$\begin{bmatrix} 1+3s^2 & 1+3s+3s^2 \\ 2+s+6s^2 & 3+7s+6s^2 \end{bmatrix}$$

It is fairly obvious that the next component is a resistor, since no frequency term is required to reduce the degree of some element; but it may be either a series resistor or a shunt resistor, R with transfer-matrix

$$\begin{bmatrix} 1 & R \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 0 \\ 1/R & 1 \end{bmatrix} \text{ respectively}$$

Since the factorization is unique,¹ only one of these is correct. Thus if one is tried and does not give a reduced matrix of the correct form, the other factor must be used. Using the series resistance-matrix in the above problem, and choosing $R = 1/2$ gives a reduced matrix:

$$\begin{bmatrix} -\frac{1}{2}s & -\frac{1}{2}-\frac{1}{2}s \\ 2+s+6s^2 & 3+7s+6s^2 \end{bmatrix}$$

This matrix requires premultiplication by a negative admittance matrix in the next stage, and hence the factorization fails. Thus the factor must be a shunt resistor, and a choice of $G = 1/R = 2$ gives

$$\begin{bmatrix} 1+3s^2 & 1+3s+3s^2 \\ s & 1+s \end{bmatrix}$$

Such a trial and error approach is unavoidable in this type of problem, but the amount of work done to detect a faulty trial can be minimized by inspecting every new coefficient as it is calculated; as soon as a negative coefficient is recognized, the factor may be assumed faulty; otherwise it is the correct factor.

The above example yields the network of Fig. 4.

6. General Passive Networks

These are characterized by the general form of A -matrix, for example:

$$A = \frac{1}{4s^2} \begin{bmatrix} 144s^2+352s^4+96s^6 & 2s+7s^3+2s^5 \\ 64s^3+96s^5 & s^2+2s^4 \end{bmatrix}$$

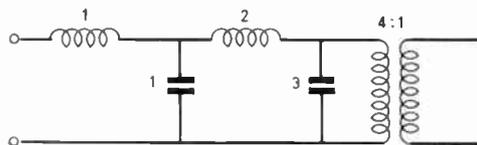


Fig. 3. Derived lossless network.

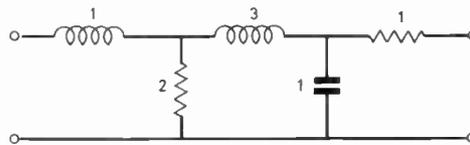


Fig. 4. Derived lossy network.

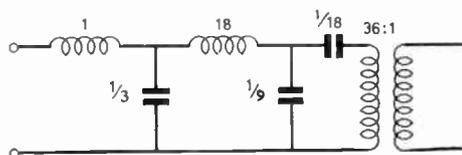


Fig. 5. One solution to the problem in Section 6.

If, as in this example, $g(s)$ is a simple multiple of s^2 , or if, as may be true for a lossy network matrix, $g(s)$ is a simple multiple of s , the problem may easily be reduced to a form similar to the simple problems considered previously. For the above example may be written as

$$\begin{bmatrix} 36+88s^2+24s^4 & \frac{1}{2}s^{-1}+\frac{7}{4}s+\frac{1}{2}s^3 \\ 16+24s^3 & \frac{1}{4}+\frac{1}{2}s^2 \end{bmatrix}$$

although $A_{12}(s)$ is no longer strictly a polynomial in s . Applying the methods of Section 4, the solution to this problem gives the network of Fig. 5 which is somewhat different to the network from which the original matrix was derived. This shows that the factorization is no longer unique; this is due to the method of eliminating $g(s)$.

The general method for this type of problem is to leave the matrix in the form

$$A = \frac{1}{g(s)} \begin{bmatrix} B_{11}(s) & B_{12}(s) \\ B_{21}(s) & B_{22}(s) \end{bmatrix}$$

and then use one of the following criteria to find the elementary matrices:

- (i) reduce the degree of some element of the matrix;
- (ii) make $g(s)$ a factor of every element of the matrix so that $g(s)$ may be eliminated.

The uniqueness of the solution is lost because (ii) may be used at several different stages of the factorization. For example,

$$\frac{1}{4s^2} \begin{bmatrix} 144s^2 + 352s^4 + 96s^6 & 2s + 7s^3 + 2s^5 \\ 64s^3 + 96s^5 & s^2 + 2s^4 \end{bmatrix}$$

Premultiplication by A_z^{-1} with $z = 2s^{-1}$ gives

$$\begin{aligned} & \frac{1}{4s^2} \begin{bmatrix} 16s^2 + 160s^4 + 96s^6 & 3s^3 + 2s^5 \\ 64s^3 + 96s^5 & s^2 + 2s^4 \end{bmatrix} \\ &= \begin{bmatrix} 4 + 40s^2 + 24s^4 & \frac{3}{4}s + \frac{1}{2}s^3 \\ 16s + 24s^3 & \frac{1}{4} + \frac{1}{2}s^2 \end{bmatrix} \end{aligned}$$

which may be factorized by the method of Section 4. The total network synthesized this time is as shown in Fig. 6 and is the same as that from which the matrix was formed. But yet another network may be synthesized by leaving the elimination of $g(s)$ until the third stage of the factorization; the network in this case is shown in Fig. 7.

In the general case, there may be several solutions to this type of problem, although eliminating the factor at some stage may result in a new matrix which cannot be reduced any further. Provided the matrix has a determinant of unity, a solution exists, and it is then a problem of finding all possible solutions and selecting the most desirable one.

7. Conclusions

Davies^{1,3} has previously shown that the factorization of certain classes of transfer matrix, by the methods described here, is unique. Above it has been shown that rules may easily be set up for such factorizations for simple problems, although trial and error methods are unavoidable when there is resistance present in the circuit. In the general problem, the solution is no longer unique, and it is then necessary to find all solutions.

The main application of the transfer matrix method of synthesis is in the design of filters; to be more precise, the form of transfer matrix considered above often represents low-pass filters. High-pass filters have transfer matrices of the same form as those above, but their elements are polynomials in the variable $p = 1/s$ and provided that this substitution is applied in the previous work, the methods still work. It is also possible, although a little more

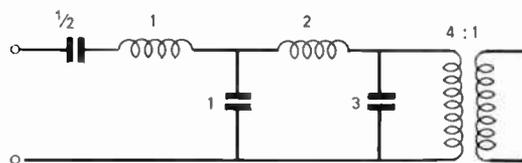


Fig. 6. An alternative solution to the problem in Section 6.

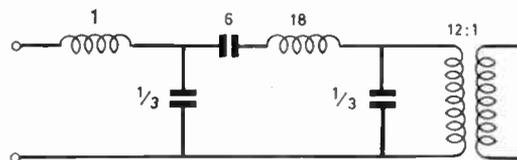


Fig. 7. A third solution to the problem in Section 6.

difficult, to apply the method to band-pass filters whose transfer matrices have elements as functions of s and p .

8. Acknowledgments

The author thanks Mr. K. G. Nichols of the Electronics Department of the University of Southampton and Dr. J. P. Cleave, late of the Computation Laboratory of the University of Southampton, for their invaluable help and guidance during the work.

The Department of Scientific and Industrial Research (now Science Research Council) provided a grant for this research.

9. References

1. R. E. Davies, 'Four-terminal network synthesis by matrix methods', *Control* (GB), **6**, pp. 100-2 (June 1963). Also 'Network Synthesis by A-matrix Factorization', B.B.C. Mathematical Memorandum No. F-1024, 1960.
2. 'Computer', 'Mathematical tools', *Electronic and Radio Engineer*, **35**, Nos. 2, 3, 4, 5 (February-May 1958).
3. R. E. Davies, 'Four-terminal network synthesis by matrix methods', *Control* (GB), **7**, pp. 21-3, July 1963. Also 'Realization of a Lossless Network in Terms of Standard 3-Terminal Networks in Cascade', B.B.C. Mathematical Memorandum No. F-1029, 1960.
4. J. E. Storer, 'Passive Network Synthesis' (McGraw-Hill, New York, 1957).
5. E. S. Kuh and D. O. Pederson, 'Principles of Circuit Synthesis' (McGraw-Hill, New York, 1959).

Manuscript first received by the Institution on 18th January 1966 and in final form on 14th October 1966 (Paper No. 1107)

© The Institution of Electronic and Radio Engineers, 1967

Electronic Instrumentation in India

A Symposium on 'Electronic Instrumentation', organized by the Central Electronics Engineering Research Institute, Pilani, Rajasthan, with the support of the Institution's Indian Division, was held on 29th April to 1st May 1966.

The following are the abstracts of some of the papers presented at the Symposium, which have now been printed in full in the *Proceedings of the Indian Division*, 4, No. 4, October-December 1966.

'Indigenous design and development of high and medium power transistorized v.h.f.-f.m. transreceivers'. P. K. RANGOLE and H. K. JAIN (*C.E.E.R.I., Pilani*)

An approach to the design and development of transistorized v.h.f.-f.m. transmitter-receivers using indigenously available components is presented in this paper. Design features and performance details of some of the basic modules of the transmitter-receivers are discussed in a general way. Comments on some incidental aspects such as standardization and economic factors are also included.

'Design of electronic circuits for the measurement of short half-lives'.

R. PRASAD and D. C. SARKAR (*Department of Physics, Muslim University, Aligarh*)

Conventional techniques used to measure the short half-life period of a radioactive isotope have certain limitations of accuracy. This paper describes the development of electronic circuits which have been designed and used for the measurement of short period radioactivities. Half-lives of the order of 0.1 second to 1 minute have been measured by this device.

'Transistorized p.h.f. magnetometer'.

M. SUDHAKARA MENON and H. N. SWAMY (*Indian Naval Physical Laboratory, Cochin*)

The smallest field that can be measured accurately by conventional magnetometers is about 2 milligauss. Proton precession magnetometers and peak height fluxgate (p.h.f.) magnetometers are commonly used to measure magnetic fields of the order of 0.1 milligauss. A peak height fluxgate magnetometer system has been developed at the Indian Naval Physical Laboratory, Cochin, using a mumetal core detector. The paper describes a transistorized peak height fluxgate magnetometer which can record a minimum field of 0.1 milligauss and the sensitivities available are 1.25 mG and 2.5 mG per inch deflection on a pen and ink recorder.

'A fast acting protection circuit for transistorized voltage regulators using series control elements'.

U. S. SINGH and G. N. ACHARYA (*C.E.E.R.I., Pilani*)

Protection of the control element against overload conditions is one of the major requirements of transistorized voltage regulators. The use of a silicon controlled rectifier (s.c.r.) offers an interesting possibility for the solution of this problem. The design of a novel protection circuit using an s.c.r. and a sensing transistor is discussed, and results obtained with a typical voltage regulator circuit using a fast acting s.c.r. protection unit are presented.

'Sub-carrier discriminators for telemetry'.

S. RAMAKRISHNA and N. V. NAYAK (*D.E.R.L., Hyderabad*)

Frequency division multiplexing systems require some form of sub-carrier discriminators to recover the data from the individual frequency channels. A four-channel f.m.-a.m. telemetry system has been designed and developed for use in rocket-borne instrumentation and for data analysis four sub-carrier discriminators have been made. An unbalanced trigger circuit type of discriminator is found to be preferable to the conventional Foster-Seely discriminator or the counting type discriminator. The channel selection is carried out by using passive filters which have been designed for the required characteristics. The output is amplified by a d.c. amplifier and directly feeds a pen-recorder. Overall performance is quite satisfactory with a dynamic input range of 30 dB and an inter-channel rejection of 40 dB. Linearity is better than 2% in the channel.

'Video tester'.

T. K. B. RAO and E. MURALI (*Defence Electronics Research Laboratory, Hyderabad*)

A video tester for tracing, checking and servicing the video stages of certain radar receivers has been developed. This unit is essentially a pulse generator giving 0.7 ms and 1.9 ms pulses. Delay line technique of pulse generation has been employed to get sharp rise time (0.06 ms) pulses. The delay lines have been designed by Guillemain's analysis. Complete circuit details are described.

Copies of the issue of the *Indian Proceedings* containing the above papers may be obtained from the Indian office of the I.E.R.E., 7 Nandidurg Road, Bangalore 6, price Rs 2 to members and Rs 5 to non-members, or from the I.E.R.E., 8-9 Bedford Square, London, W.C.1, price 3s. to members 7s. 6d. to non-members.