

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 36 No. 1

JULY 1968

Papers for the Practising Designer

THE flood of technical papers which surrounds the engineer today is overwhelming. He is continually faced with the problem of how best to select, for his own specialized and general needs, good quality papers of the right kind. Choosing at random, hoarding everything that comes to hand, or shunning everything until a specific need arises, are unscientific methods; scanning abstracts or using S.D.I. retrieval methods are far more valuable guides to world literature. Whichever way is adopted, most members of the I.E.R.E. place some reliance on the regular issues of this *Journal* and in consequence responsibility for its papers' content is exacting. To meet this responsibility the aim of the Papers Committee is to select and publish over the year a balanced though necessarily limited collection of good, interesting and suitable papers.

Now we all think we know what constitutes a good paper, but which are the suitable ones? The present high standard of the *Journal* has been built up by concentrating on certain types of paper: briefly, these are original research papers, co-ordinating surveys, and engineering development papers. Some are theoretical, others eminently practical and the annual Premiums and Awards usually reveal a balance between the best in these categories. All should rely in part on the use of critical scientific reasoning and judgement. Other categories, not necessarily any less demanding in these respects, are the tutorial type of paper, short Contributions, Letters to the Editor, and Discussion reports. The Contribution category embraces interesting and useful topics, maybe descriptive, very brief or urgent, but not necessarily original, in either experiment, theory or technology.

The Papers Committee now advocates further types of paper. Elsewhere in this *Journal* are examples of two new types of paper specially written by Dr. E. A. Faulkner of Reading University. The first, which may be described as an 'interpretive' paper, can be regarded as the near-prototype of an especially timely and valuable addition to the present range of types. Such papers 'interpret' basic theory in a context relevant to some particular purpose and should have an immediate interest and value for designers. Emphasis is on evolving the engineering design criteria and not on the subsequent details of device or system realization, manufacture or operation. The second short contribution presents design data which are not readily available to the designer. The literature is often denied both such useful types of paper due to commercial security. Again, critical engineering judgement must be seen to operate. The informal style adopted in the example quoted is deliberate. The reaction of readers will be of interest and letters to the Editor on the subject are invited. Among other new types of paper being considered is one with graduate-level students particularly in mind. For an example, a forthcoming issue of the *Proceedings* will include such a paper on Satellite Communications Systems by Mr. C. F. Davidson.

If you feel you can write such new papers, you are cordially invited to submit them to be considered for publication. The call for the other types of paper is, of course, as great as ever if a balanced content is to be maintained. The main supplier of good papers remains the individual volunteer: may his art prosper! Other sources are the Specialized Group and other Standing Committees of the Institution, presentations at evening meetings (including those of Local Sections), and the Institution's various Conventions, Conferences and Symposia. From time to time, papers are invited from eminent specialists.

You may think that it is timely to have a paper on a particular subject. If so, you are probably the best person to write it, or you probably know who could be persuaded to do so. There is not a remote and mysterious group of specialist engineers who produce papers and give lectures: they are produced by typical engineers writing on their professional activity.

E. ROBINSON

INSTITUTION NOTICES

Professor Leslie Kay

The President and several senior members of the Institution's Council recently had discussions in London with Professor Leslie Kay, Ph.D., C.Eng., F.I.E.E., F.I.E.R.E., who is Head of the Department of Electrical Engineering at the University of Christchurch, New Zealand. These discussions centred round the expansion of I.E.R.E. activities in New Zealand where there are flourishing Local Sections in Wellington, Auckland and Christchurch. Professor Kay was able to provide further information to that obtained by the Director of the Institution, Mr. Graham D. Clifford, when he visited New Zealand two years ago, on the setting up of a Division.

Members will be aware already of Professor Kay's work in connection with the ultrasonic aid for the blind which he invented when he was on the staff of the Electronic and Electrical Engineering Department of the University of Birmingham and later at Lanchester College of Technology, Coventry. He is continuing the development of this instrument and while in Great Britain he had discussions with the various organizations for the blind as well as with universities and medical institutions. On his return journey to New Zealand Professor Kay visited several leading workers in this and associated fields in the United States. His world tour, which was partly sponsored by the National Electronics Council of New Zealand, included visits on behalf of the Council to organizations concerned with sonar techniques for fish finding.

Computer Aided Design

The scope of the Conference on Computer Aided Design, announced in the May issue of *The Radio and Electronic Engineer*, has now been defined. The Conference is to be held at the University of Southampton from 15th to 18th April 1969.

All aspects of engineer-orientated computer programs and topics relevant to computer aided design and design automation hardware, software, application programs and other problems will be covered. On-line applications, whereby there is conversation between engineer and computer during the design process, will be of particular interest and will include the application of techniques such as graphics, information retrieval, and man-machine communications. Off-line engineer-orientated programs of a general nature that might at a later date be modified for on-line operation, will also be dealt with.

Contributions of up to 3000 words are invited for consideration in the programme. Completed texts

will be required by 15th November 1968, but a 250-word synopsis should be submitted to the Conference Secretariat at the I.E.E., before 30th August 1968.

An exhibition of apparatus is planned and suggestions for this, which should be accompanied by a technical description of the apparatus concerned and details of the size and power requirements, will be welcomed.

The Conference is being held under the aegis of the United Kingdom Automation Council, and is sponsored by the I.E.E., the I.E.R.E., the I.E.E.E. (U.K. and Republic of Ireland Section), the I.Mech.E. and the University of Southampton. Further details may be obtained from the Conference Secretariat, Institution of Electrical Engineers, Savoy Place, London, W.C.2, or from the I.E.R.E., 9 Bedford Square, London, W.C.1.

Electronic Engineering in Oceanography

The Revised Proceedings of the Conference on Electronic Engineering in Oceanography, which was held in Southampton in September 1966, is now published. Orders may be placed with the Publications Sales Department of the I.E.R.E. and should be accompanied by a remittance of £6.

All who registered to attend this Conference, and any other persons who purchased the original Proceedings, will receive a Supplement Volume containing additional papers, discussion reports, corrigenda and lists of delegates. Supplements will be sent out as soon as possible.

Correction

Due to a printing error four lines were transposed in the text of the paper 'Propagation of uniform plane waves in a Fabry-Perot Resonator containing Ammonium Dihydrogen Phosphate (ADP) as Dielectric' (*The Radio and Electronic Engineer*, June 1968).

Page 347, column 2, Section 4.2. should end at five lines after eqn. (13). Section 5. 'Application to Fabry-Perot Resonator with ADP as Dielectric', should start with the sentence 'As in the case of a resonator . . .'

Index to Volume 35

The June issue completed Volume 35 of *The Radio and Electronic Engineer*. The Index for January to June 1968 will be included in the August issue of the *Journal*.

Radar Pulse-compression by Random Phase-coding

By

G. C. BAGLEY,
C.Eng., M.I.E.R.E.†

Summary: Integrated-circuit binary shift registers are used to realize the matched-filter to randomly coded radar signals. The system exploits the favourable range side-lobe properties of randomly changing binary codes to achieve pulse compression ratios greater than those obtainable with fixed coding schemes. The output noise distribution, signal/noise ratio, range resolution characteristics, range side-lobes and range ambiguities, and Doppler response of such a radar system are discussed.

List of Symbols

n	number of digits in binary code
P_e	probability of a decision error
P_c	probability of a correct decision
E	average signal energy per sub-pulse (watt-seconds)
N_0	average noise power per unit bandwidth (watts per hertz)
ρ	the correlation between the binary waveforms
σ	standard deviation (noise voltage)
σ^2	variance (noise power)
θ	signal phase angle
ϕ	output phase angle

1. Introduction

Pulse compression is a technique employed in the field of pulsed radar to increase the range resolution by the use of a special modulation within the radar pulse. In the receiver of a pulse compression radar, a cross-correlation process between the received signal and a stored replica of the transmitted signal is effected in a device termed a matched-filter.¹ The function of the matched-filter is to maximize the response of the system to the signal to which the filter is matched, whose stored replica it holds as a pattern of the expected signal. The conditions which must be satisfied in order that the filter may be matched to the signal may be expressed equivalently as:

(1) The impulse response $h(t)$ of the filter must be the time reverse of the signal waveform $s(t)$. Ignoring any constant delay necessary for realizability,

$$h(t) = s(-t).$$

(2) The complex frequency response $H(f)$ of the filter must be the conjugate of the spectrum $S(f)$ of

the signal, that is, $H(f) = S^*(f)$. $H(f)$ with $h(t)$, and $S(f)$ with $s(t)$ are Fourier transform pairs.¹

When the matched signal is applied at the input of the filter, the output is the auto-correlation function (a.c.f.) of the signal waveform. It is the central 'spike' of this from which the compressed pulse is derived. When, as in practice, random receiver noise is also present, the signal/noise ratio at the output of a matched-filter has the maximum value obtainable. The noise, being random and unpredictable, receives no special favour, whilst the matched signal which is well known to the filter, is enhanced relative to the noise by the so-called pulse compression ratio. This is equal to the time-bandwidth product of the signal.

One typical pulse compression waveform is a pulse of radio frequency with a linearly swept frequency and having therefore a quadratic phase characteristic. To take another example, if the phase were restricted to discrete values, rather than the continuum just cited, then the phase values might be determined by the elements of some code sequence, and we then have 'digital pulse compression'. A suitable matched filter might then take the form of a tapped delay line, and suitable weighting is then applied to the outputs of the several taps before they are coherently added.

Tapped delay lines are critical of adjustment and, because of their bulky nature, limit the number of digits in the code. When a fixed code pattern is used, as dictated by a fixed pattern of weightings, pulse compression ratios are further limited by the range side-lobes which accompany the compressed pulse. These are the outer fringes of the temporal a.c.f. of the signal and become troublesome when fixed code sequences longer than 13 digits are used. Side-lobes of strong signals appear as spurious returns time related to the compressed pulse, and may interfere with weaker returns of different range.

In the present paper inexpensive and compact integrated circuit binary shift-register stages replace

† Ministry of Technology, Royal Radar Establishment, Malvern, Worcestershire.

the expensive and cumbersome passive delay line, and the generation and processing of relatively long coded sequences become possible.

The problem of the range side-lobes is solved by means of a random coding technique whereby the transmitted code sequence is changed from pulse to pulse, and is stored in a memory device to enable the receiver to be correctly matched to the signal. The nature of the resulting random side-lobes is such that they resemble the response of the system to receiver noise since they occur randomly in both amplitude and position. This eliminates the repetitious false-alarms associated with the side-lobes of a fixed code sequence.

2. The Digital Matched-filter

It is well known that a filter having an arbitrary impulse response may be synthesized by the summation of the weighted outputs of a tapped delay line. The tapping points are usually equally spaced, and various phase shifts may be applied to the outputs to ensure coherent addition (Fig. 1). In the digital matched-filter (d.m.f.) a binary shift register, driven by suitable clock pulses, is used in the role of the delay line, and the weighting coefficients, which for

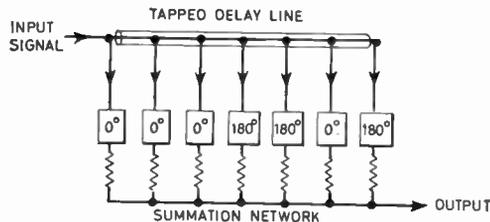


Fig. 1. Passive delay-line matched-filter.

simplicity are restricted to +1 and -1, are obtained by taking the erect or inverted outputs of the shift-register stages. Because each digit of the sequence occupies each stage for the duration of one clock period, this corresponds to the temporal separation between the tapping points of the equivalent passive line and to the sub-pulse duration of the coded signal. When the code sequence to which the filter is to be matched is fixed, the pattern of the tapping connections is the reverse of the sequence. This is necessary because the early digits of the code are the first to reach the distant end of the shift register (Fig. 2). The result of this is that the response of the filter to a single impulse, equal in duration to one clock-period, is the time-reverse of the matched sequence. This satisfies one definition of a matched-filter.² The technique can be used to match any chosen binary sequence (Fig. 3).

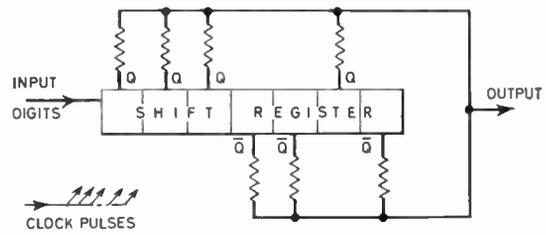


Fig. 2. Digital matched-filter to a fixed code sequence (0100111).

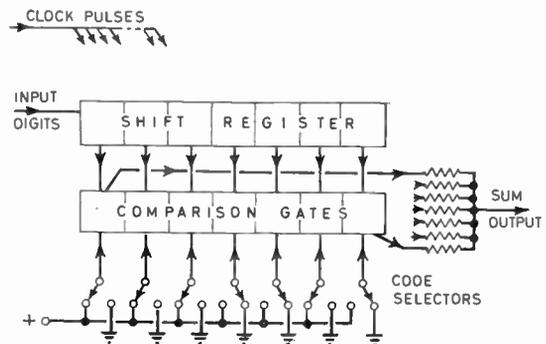


Fig. 3. A digital matched-filter to any chosen seven digit code.

When the matched sequence is changed on a pulse-to-pulse basis, a stored replica of each transmitted sequence must be used to set the appropriate pattern of erect and inverted outputs from the various stages of the shift register. This may be effected by means of a bank of equivalence gates, arranged to compare the state of each stage of the register with the stored code digit in the code memory. The stored code sets the shift register filter to respond to each sequence and is changed at each transmission. Where there is agreement between a received and a stored digit, a unit contribution is made via each input resistor of a summing network (Fig. 4). A shift register of n stages will match an n -digit sequence, and the filter will generate $n + 1$ uniformly distributed output voltage levels. Complete agreement between the digits occupying the n stages of the register and the n digits of the stored code will result in n contributions to the output summation network, giving a compressed pulse of maximum amplitude, and equal in duration to the clock-period.

In the case of both fixed and changing codes, a sequence of inverted sign, such as that which occurs when the input signal is phase inverted, will produce a compressed pulse of inverted polarity since there will then be n stages having full negative correlation with the stored sequence.

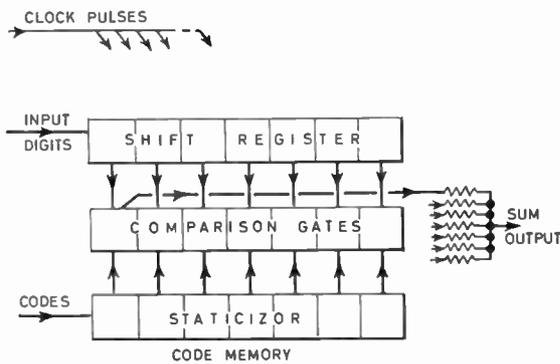


Fig. 4. Digital matched-filter to a randomly-changing code.

The response of the digital matched-filter to its matched sequence is the a.c.f. of that sequence for discrete increments of time equal to the clock-period. Thus the early digits of the code are firstly correlated with the code settings of the early stages of the register, representing the tail end of the code, and then, after the occurrence of the compressed pulse, the tail end of the sequence, being last to leave, is correlated with the pattern of the early part of the code associated with the final stages of the register.

The auto-correlation function of the code may be expressed as

$$C_k = \sum_{i=1}^{n-k} a_i \cdot a_{i+k} \dots\dots(1)$$

where a_i, a_{i+k} are the elements of the code. They have the values +1 or -1.

The correlation effect of the shift register is illustrated in Fig. 5. In this diagram, the binary sequence from the receiver is seen to move (from left to right) along the shift-register under the control of the system's clock pulses. At each step, the code settings a_k indicated at the bottom of the diagram, are compared digit by digit, with those digits occupying the shift register. The outputs of those gates where agreement occurs (a pair of '1's, or a pair of '0's) cause a positive increment to be made via one of the inputs

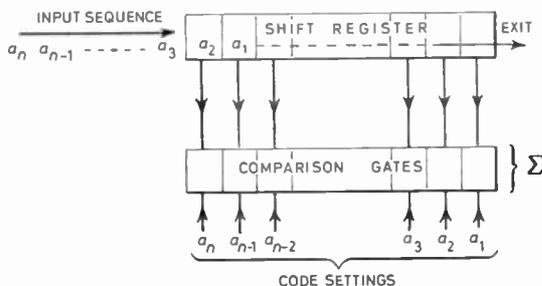


Fig. 5. Correlation process in the digital matched-filter.

of an analogue summation network. A negative increment results if the stored and received digits differ.

In a binary system a code sequence cannot exist in isolation, and must be processed in a context of preceding and following random digits corresponding to quantized receiver noise. The noise digits are equally probable '0's and '1's, and there is no other alternative state for the shift register, since the system is binary.

In the laboratory, it is possible, by constructing a pair of channels in parallel, to study the separate responses of the systems to both the noise digits and the code digits. This is done by inverting the noise context in one channel relative to the other, the embedded code sequence being erect in both channels (Fig. 6). The matched-filter outputs may then be added or subtracted in order to display either the signal or the noise response.

The output of each matched-filter is the sum of n separate contributions from the stages of the shift register. When these are random, the output voltage levels have a (discrete) binomial distribution of order n , i.e. $(n+1)$ levels. When the noise digits are displaced from the register by the entry of a code sequence, the order of the binomial distribution due to the remaining noise digits is reduced to consecutive lower orders (alternately odd and even) as each code digit enters the register, and then, reaching zero (when the code is in complete possession) increases one order at a time until the complete noise distribution of order n is restored. This is illustrated in the bottom trace of Fig. 7.

For the random code sequences the converse takes place, resulting in binomial distributions of orders from unity, when only the first digit has entered, up to the $(n-1)$ th order. Then, for the matched sequences, the compressed pulse occurs, followed by decreasing distributions from order $(n-1)$ down to

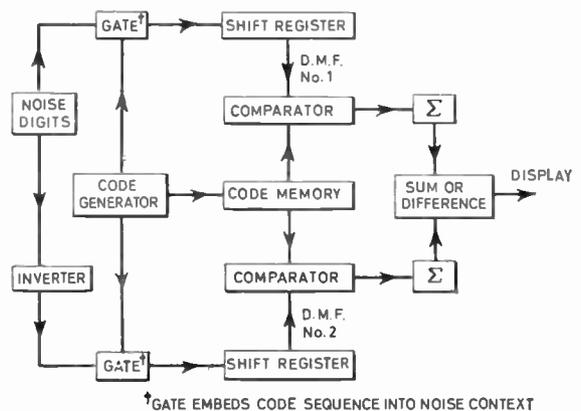


Fig. 6. Test arrangement for code evaluation.

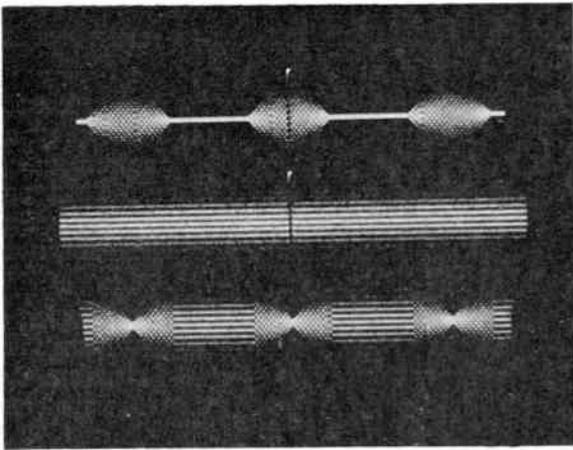


Fig. 7.

- Top trace: Response of digital matched-filter to uncorrelated, correlated, and again uncorrelated sequences.
- Centre trace: The compressed pulse of top trace in context of noise.
- Bottom trace: The separate response of the filter to the noise context removed from the top trace.

unity. Since it will be recalled that the output of the digital matched-filter to its matched sequence is the a.c.f. of the sequence, the above family of output distributions represents the superimposed a.c.f.s of a number of random sequences, and the response of the filter when it is matched to each one in turn is shown in Fig. 7 (top centre trace). Each a.c.f. has identical but reversed structure on either side of the compressed pulse. The wedge-shaped envelope is the a.c.f. of the envelope of the uncoded long pulse of n identical digits. In Fig. 7 the outer top trace represents the superimposition of a number of cross-correlation functions between random sequences when the filter settings are not matched to these particular sequences. Again there is a wedge-shaped envelope, but no compressed pulse, and therefore no range ambiguity. The unmatched response of the filter to the context of random digits due to noise has a corresponding wedge-shaped gap into which the random a.c.f.s of the code sequence fit. When the two responses are combined by the use of only one matched filter, an undisturbed series of symmetrical distributions of order n results. Compressed pulses rise out of this noise background as shown in the centre trace of Fig. 7. The oscilloscope traces in Fig. 8 show the corresponding situation when the code sequence does not change. On the top trace of Fig. 8, the same a.c.f. from each pulse is shown. The centre trace shows the resulting range ambiguities. The last trace of Fig. 8 refers to the noise content, and is similar to that of Fig. 7.

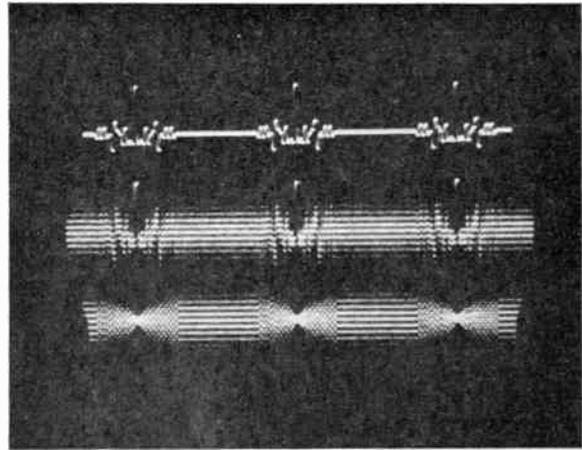


Fig. 8. As for Fig. 7, but with fixed coding.

3. The Transmitted Signal

The radar pulse is sub-divided into a number (n) of discrete contiguous sub-pulses, identifiable as a pattern of phase angles within the long pulse, and known *a priori* to the receiver. The receiver is supplied with information concerning the fine structure of the signal once per pulse repetition interval. The phase modulation of the carrier takes the form of 180° phase reversals in accordance with the binary digits of each code. This gives maximum negative correlation between the binary elements of the signal, and makes most efficient use of the transmitted energy. It will be seen in the discussion of error probabilities (Sect. 5) that a correlation of -1 between the '0' and '1' sub-pulses gives minimum decision error in the receiver. When, as in the present case, the digits are equiprobable, and therefore on the average equal in number, phase reversal keying resembles double side-band suppressed-carrier modulation in that there is maximum energy in the side-bands, i.e. there is no d.c. term in the modulating function. It is the modulation by the narrow sub-pulses that gives the signal a wide bandwidth, and thereby high range-resolution.

Phase modulation by higher, for example quaternary, digital codes is also possible, but while the simple binary codes lend themselves to implementation by inexpensive digital integrated circuits, processing polyphase signals seems to be best suited to passive matched filters.

The binary phase modulation may be applied to the radar signal at almost any point in the transmitter chain. When the digit rate is high, the overall bandwidth is greater, and frequency stability is not a serious problem. However, the use of longer overall pulse durations places some restrictions on frequency stability, and also limits the range of Doppler shifts to be tolerated by the matched-filter system. The

latter is also a function of which radar frequency band is to be used, and is discussed in Section 11.

The selection of random codes transmitted need only be random over a time which exceeds the integration time employed in the receiver.

4. The Detection System

Since, in general, a target has unknown range, the phase angle of the received signal is unknown. It is convenient to resolve the signal into its quadrature components, and to process these in separate channels. These may be termed *I*, or the in-phase, and *Q*, or the quadrature channels. The incoming signal may therefore be demodulated by a pair of phase-sensitive detectors having reference voltages at the carrier frequency of the i.f. and in phase-quadrature. Each detector is provided with an integrator, or low-pass filter having a $(\sin x/x)$ type of frequency response. This is equivalent to having a rectangular impulse-response, and therefore provides a matched filter to the individual sub-pulses of the coded long pulse.³ The output voltage will be of opposite polarity for the two binary phase angles, and a sampling circuit, or analogue-to-digital converter (a.d.c.), then converts this to binary digits at the clock frequency. (The timing of the sampling pulses is discussed separately in Section 8.)

When no identifiable signal is present, the a.d.c. makes random decisions based on the average receiver noise at the sampling intervals. The receiver noise, assumed to be Gaussian, is resolved into two independent orthogonal components by the phase-sensitive detectors, which, when quantized by the sampling process, give rise to a pair of 'coin-toss' binary waveforms. If the two reference voltages of the detectors are truly orthogonal, these two random sequences will be statistically independent and each digit may be, with equal probability, (=0.5) either 0 or 1.

When the phase-coded signal is received, the sub-pulses will also be resolved into their two components, which are quantized by their respective a.d.c.s, and supplied to the digital matched-filters as coded sequences. These may each be inverted or erect, depending upon the quadrant of the Argand diagram into which the signal vector falls. This is a function of target range. A digital matched-filter in each quadrature channel produces a short 'compressed' pulse, equal in duration to the clock-period, each time a coded sequence is successfully recognized by the receiver. Finally, as in the classical two-path detector system, the two channels are recombined by squaring and adding to produce a unipolar output pulse. A block diagram giving details of the digital processor is shown in Fig. 11.

A detection threshold device may, if desired, be employed after recombination. It has been shown that this is preferable to the use of separate thresholds within the quadrature paths before combination.⁸ The combination of post-demodulation low-pass filtering and binary quantization of the separate quadrature components gives the system some of the properties of a band-pass limiter. This affects the determination of the signal/noise ratio (s.n.r.) which is discussed in Section 7.

5. Decision Errors

In the presence of noise, an element of doubt exists as to the true identity of each binary digit. The error probability, P_e , ranges from zero, for a very strong (erect) signal, to 0.5 for the almost blind guess made by the decision circuit for a very weak signal. The probability of error is a function of various factors, which include signal/noise ratio, signal phase-angle, the nature of the binary wave forms, and the degree of synchronism between the received code elements and the clock intervals of the sampling process. The latter calls for special consideration and is discussed later.

Several authors have discussed the probability of errors in the demodulation of digitally phase-modulated signals. The most useful expression is that derived from basic energy considerations by Lawton:⁴

$$P_e = \frac{1}{2} \left(1 - \operatorname{erf} \sqrt{\frac{E(1-\rho)}{2N_0}} \right) \quad \dots\dots(2)$$

where E = average energy per sub-pulse (joules)

N_0 = average noise-power density per unit bandwidth (watts/Hz),

ρ is the cross-correlation coefficient between the two binary waveforms,

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du \quad (\text{'error function'})$$

For binary phase modulation, $\rho = -1$, and the expression simplifies to

$$P_e = \frac{1}{2} (1 - \operatorname{erf} \sqrt{E/N_0}) \quad \dots\dots(3)$$

The term E/N_0 is an energy s.n.r., and $\sqrt{E/N_0}$ is therefore proportional to signal amplitude, and may be corrected for the phase error θ between the modulated signal and the reference voltage of the phase-sensitive detectors. Montgomery⁶ gives a similar expression in his equation (14):

$$P_e = \frac{1}{2} \left(1 - \operatorname{erf} \left(\sqrt{\frac{E}{N_0}} \cos \theta \right) \right) \quad \dots\dots(4)$$

If, due to target range, a signal having inverted phase is received, the cosine term is negative. Unlike the telegraphic case, the digital matched-filter of a radar has no preference for erect or inverted sequences, and a negative-going compressed pulse results from a phase-inverted signal.

When the signal/noise ratio is small, $\sqrt{(E/N_0)}$ is small, and if there is no phase error, $\text{erf} \sqrt{(E/N_0)}$ tends to $\frac{2}{\sqrt{\pi}} \sqrt{E/N_0}$.

In this case,

$$P_e \approx \frac{1}{2} \left(1 - \frac{2}{\sqrt{\pi}} \sqrt{E/N_0} \right) \dots\dots(5)$$

In eqn. (5), the coefficient $2/\sqrt{\pi}$ gives rise to a slightly optimistic value (1.06 dB) of effective input s.n.r. This may be explained when the filtering and sampling operations are regarded as a band-pass limiter.

$$(S/N)_{\text{out}} = (S/N)_{\text{in}} \left[\frac{1 + 2(S/N)_{\text{in}}}{4/\pi + (S/N)_{\text{in}}} \right]$$

taken from Gardner⁹ based on the work of Davenport.⁷ This shows that weak signals incur a loss of $\pi/4$ which appears within the square-root sign of equation (5) and cancels the $2/\sqrt{\pi}$ outside. This gives, for small signals,

$$P_e = \frac{1}{2} (1 - \sqrt{E/N_0}) \dots\dots(6)$$

or

$$P_e = \frac{1}{2} (1 + \sqrt{E/N_0}) \dots\dots(6a)$$

The weak signal case is of considerable importance in the phase-coded radar case. When high pulse-compression ratios are sought, the total signal energy must be divided between the many sub-pulses

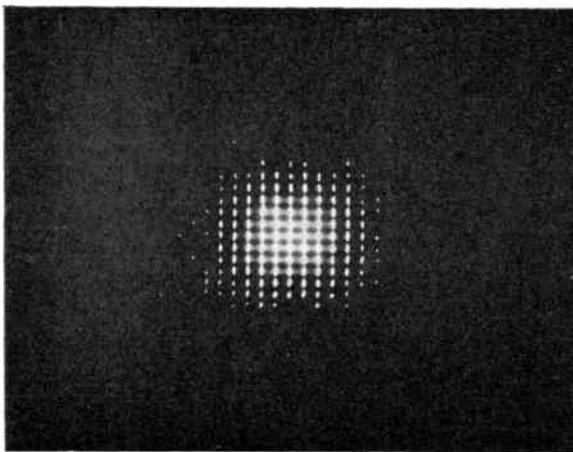


Fig. 9. The (no-signal) bivariate binomial-distribution due to quantized receiver noise.

of the signal, and individual errors will be numerous. It is only for very strong signals that nearly correct sequences may be expected, and sequences containing many errors may be processed in the system described.

6. Output Noise Distribution

When only random input digits are present, the output noise of each d.m.f. has a symmetrical binomial distribution which, when n is large, approximates to

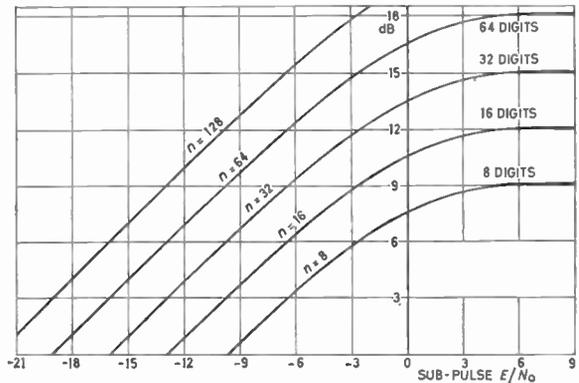


Fig. 10. Output s.n.r. plotted against input (sub-pulse) s.n.r. (dB).

the normal distribution. The mean output noise voltage for a single d.m.f. is the standard deviation σ , of this distribution. This is the square-root of the mean noise power σ^2 given by $\sigma^2 = np(1-p)$, where p is the probability of a '1', and $1-p$ the probability of a '0' (see Ref. 5). For receiver noise, $p = 0.5$, $\sigma^2 = n/4$ and $\sigma = (\sqrt{n})/2$. The unit of voltage is the 'filter-output-voltage-level'. (This is about $1/n$ of the supply-rail voltage.)

When a pair of quadrature channels is employed, two such output distributions are produced, each independent, and corresponding to the two orthogonal components of the Gaussian receiver noise. The outputs of the two quadrature channels may be combined by squaring and adding, and possibly subsequently taking the square root of the resulting voltage. An insight into the output voltage distribution may be obtained if the matched-filter outputs are applied directly to the vertical and horizontal inputs of an oscilloscope, when the resulting bivariate distribution will be displayed. For receiver noise, the digit probabilities are equal, and for $n = 16$ the distribution is shown in Fig. 9. When a signal is present, the values of P_e for the I and Q channels differ, as they are dependent on the resolved components of the signal. When the resulting sequences are in complete possession of the digital matched-filters, the resulting radial deflection of the peak of

the bivariate distribution will have a direction determined by the signal phase angle, a function of target range.

For the *I* channel

$$P_e = \frac{1}{2}(1 - \text{erf}(\sqrt{E/N_0} \cos \theta)) \quad \dots\dots(7)$$

For the *Q* channel

$$P_e = \frac{1}{2}(1 - \text{erf}(\sqrt{E/N_0} \sin \theta)) \quad \dots\dots(8)$$

where θ is the phase angle of the signal.

The modulus of the resulting radial displacement is the combined output of the two channels of the receiver, and increases with signal strength. For weak signals, this is a linear function of input signal/noise ratio. When the signal is of medium strength, the shape of the error function causes the signal amplitude to become compressed. The degree of compression depends on where in each quadrant of phase θ lies, but for practical radar returns the weak signal, quasi-linear mode may be expected. In the diagram (Fig. 10) output s.n.r. is indicated as a function of input s.n.r. (for one quadrature component only).

It is to be noted that the phase angle of the resulting displaced output is not a linear function of input phase, since the resolved components of the input signal are independently 'compressed' by the error-function characteristic.

Thus the output signal voltages of the two quadrature channels are given by

$$V_I = \frac{n}{2} \text{erf}\left(\sqrt{\frac{E}{N_0}} \cos \theta\right) \quad \dots\dots(9a)$$

and

$$V_Q = \frac{n}{2} \text{erf}\left(\sqrt{\frac{E}{N_0}} \sin \theta\right) \quad \dots\dots(9b)$$

The output phase angle ϕ , i.e. the angle at which the output probability distribution is deflected, is given by

$$\tan \phi = V_Q/V_I \quad \dots\dots(10)$$

which is approximately equal to the input phase angle θ for weak signals, and lies in the same quadrant. A moving target will cause the vector of displacement to rotate at the Doppler frequency. As signal strength increases, the vector dwells less at the centres of the sides of the distribution, and progressively longer at the corners. The effect is that the compressed pulse, when present, has a preference for the corners of the distribution, where it happens to be $\sqrt{2}$ times greater than at the middles of the sides. This fact affects the output dynamic range and is reconsidered later.

7. The Calculation of Signal/Noise Ratio

There are two cases to be considered in the calculation of the output signal/noise ratio. Firstly, in the case of very strong signals, the correct identifica-

tion of the code digits is almost a certainty. In each of the quadrature channels, the effect of correctly identified digits is to displace the first moment of the output voltage distribution, which becomes asymmetrical. The output signal amplitude is the magnitude of this displacement, and a limit is reached when all the digits have been correctly identified. The signal amplitude is then $\pm n/2$ voltage levels; the noise is $(\sqrt{n})/2$ voltage levels, and the maximum value of s.n.r. is $\sqrt{n} : 1$ in a single channel.

When the signal is weak, however, the decision circuit often guesses wrongly. The probability of being correct P_c is slightly greater than 0.5, and is given by eqn. (6a). This value may be used to compute the displacement of the output distribution, which now becomes a skew binomial distribution.

In the Appendix, this method is used to obtain an expression for the output s.n.r. of the digital matched-filter when, due to noise induced decision errors, a partially incorrect sequence is processed. It is given by $(\sqrt{n}) \text{erf} \sqrt{(E/N_0)}$, and has the maximum value \sqrt{n} already quoted.

Table 1 shows typical cases of input and output signal/noise ratios, and indicates the extent to which the output s.n.r. is affected when the correction for band-pass limiting is applied.

Table 1

Input and output signal/noise ratios for codes of 16 and 32 digits

Code length (digits)	Input s.n.r. (dB)	Output s.n.r. (dB)	
		uncorrected	corrected
16	0	0.87	0.15
16	3.01	3.7	3.25
16	6.02	6.37	6.31
16	9.03	8.73	9.09
16	12.04	11.55	11.09
16	15.05	11.64	11.92
32	0	0.96	0.08
32	3.01	3.88	3.16
32	6.02	6.71	6.26
32	9.03	9.38	9.32
32	12.04	11.74	12.10
32	15.05	13.56	14.10

For weak signals, where $\text{erf} \sqrt{(E/N_0)}$ is directly proportional to $\sqrt{(E/N_0)}$, the output s.n.r. depends little upon the number of digits in the code, but rather upon the total energy in the long pulse. Increasing the number of digits in the long pulse would subdivide the energy between the sub-pulses, giving an increased error rate, but the effect would be almost exactly

off-set by their increased numbers, and the more compact peak of the probability distribution due to noise. It is an apparent paradox of this system that in spite of the digitization into only two states, there is a linear relationship, for weak signals, between the input and output signal/noise ratios. This is because information regarding input signal amplitude is implicit in the correctness of the quantized digits (P_c), and is recovered by their summation in the output of the digital matched-filter. For code lengths in excess of 16 digits, and for signal/noise ratios up to approximately 10 dB, the departure from the quasi-linear mode is only slight.

When signals are strong, the output s.n.r. may only be increased by the use of a longer code, because the higher order binomial distribution has a more compact peak, and the output noise level (σ) is reduced relative to the amplitude of the compressed pulse. However, with codes of only 16 digits, the maximum value of output signal/noise ratio would seem to be already adequate (Table 1).

8. Sampling-time Errors

Matched filters may be either 'time varying' or 'time invariant'. The first kind are matched to the signal only at one or at several discrete points in time, and usually find application, for example, in tracking operations. The 'time-invariant' type is matched to the correct waveform regardless of when it occurs.

In the field of telegraphy, the term 'inter-symbol interference' is used to describe the influence of a particular message element or digit upon its neighbours. This type of interaction between adjacent sub-pulses may be present when, due to target range, the sampling times of the analogue-to-digital converters do not coincide exactly with the phase-transitions of the sub-pulses. Two effects contribute to the errors resulting from this mismatch. Firstly, a decision may be made as to the identity of a particular digit before all the energy of the sub-pulse has been allowed to make its contribution. Secondly, some energy belonging to an adjacent sub-pulse is included, and whose undesired contribution may either randomly reinforce or partially inhibit the correct decision in the same way that receiver noise does.

When, in the worst case, the sampling intervals lie half-way between the signal phase-transitions, it is impossible for the resulting binary decision to favour the sequence representing one range resolution cell more than its neighbour, because the binary sequence cannot simultaneously represent both a code and its time-displaced version. When, due to varying target range, further time displacement takes place, the attention of the whole system is shifted to the next

range resolution cell, which results in the appropriate independent compressed pulse.

The whole effect calls for special treatment in the receiver in a manner analogous to the provision of separate quadrature phase-reference voltages in the detector stages in order to cater for various target phase angles. Indeed, the interleaving of the sampling pulses of the two channels in 'time quadrature' is one possible solution. Another solution would be to make the clock pulses of both channels early and late by half a period on alternate scans.

The response of the complete matched-filter system is therefore time varying to the signal sub-elements, but the filter is always correctly matched to the code pattern, if not the signal, regardless of when it is received. The effect to be expected when the bandwidth of the system is increased to the point where sub-pulses are comparable in length to targets is a matter for conjecture, depending on the geometry of the targets.

9. Signal Energy and the Digital Matched-filter

There is an important fundamental difference between a conventional matched-filter, in which the spectral components are coherently added, and the digital matched-filter, in which code agreements are counted. It might be argued that the latter is not a matched filter in the true sense at all, but some sort of pattern-recognizing artifice.

However, the matching in time to a specified sequence (rather than matching a distribution in the frequency domain) also implies matching to the spectral distribution of the signal to which the time domain is connected by a Fourier-transform relationship. To the extent that the code digits are correctly identified, the filter matches the spectrum of their modulated waveform.

The magnitude of the response is seen to depend on the number of correct digits guessed which is in turn a function of s.n.r. and therefore a function of the signal energy. Thus, although it is not possible, as in the analogue matched-filter, to bring spectral energy 'into step', the energy of the received signal is implicit in the correctness of the binary decision which each digit represents. In effect, it is the information regarding the signal, rather than the signal itself, that is being processed. Thus, although this is a hard-limiting binary system, information regarding signal strength is not, in general discarded.

The possibility of a two-level processor having an extra shift-register to enable both weak and strong sub-pulses to be distinguished has been considered, but was discarded because the response to weak signals is already a linear one with an amplitude resolution determined by the number of digits in the code.

10. The Ratio of the Compressed Pulse to Mean Side-lobes

Since, for random coding, the side-lobes are indistinguishable from the response to noise, the maximum signal to mean side-lobe ratio is equal to the maximum output signal/noise ratio, which is shown to be $\sqrt{n} : 1$.

Since the side-lobes are random, bipolar, and indistinguishable from receiver noise, further post-detection integration will enhance the signal relative to both the output noise level and to the side-lobes. This was the feature which suggested binary random coding.

10.1. Uncorrelated Pulses

The code sequences are changed from pulse to pulse, and being random, a given pulse will be modulated by a code which is uncorrelated both with its predecessor and its successor. Thus the setting of the store which is matched to a given pulse will be mismatched to the previous, and next pulses. Figure 7 (top trace) shows the response of the filter to three such pulses. The noise response has been subtracted (as described elsewhere) in order to display the auto-correlated (top centre) codes, and the cross-correlated random codes on the same trace.

The foregoing discussion implies that the radar p.r.f. may be increased above the value which would normally give rise to range ambiguities, and yet ambiguity does not arise since the response of the matched filters to all sequences other than those to which they are matched is noise-like. The centre trace of Fig. 7 will illustrate this by comparison with the top trace. This is not to say that the ranges at which the cross-correlated responses occur may be ignored. Masking of other weak targets may still take place where an undesired uncorrelated return coincides with a weak wanted return, appearing as a competing noise. This effect would be most severe in extended clutter, reducing the advantage to be gained by an increased p.r.f. The desired targets will however appear through such clutter whenever there is a substantial 90° component in their relative phase angle, or when their velocities differ. The technique can, however, provide a means of obtaining increased mean power in the transmitted signal, without ambiguous indications of range. A memory staticizer must be provided at the digital matched-filter for each code sequence which is 'in flight' at any time.

10.2. Dynamic Range

The output dynamic range of the receiver is determined by the number of stages, n , in the digital matched filters. The upper limit is fixed by the ratio

of the signal at the corners of the output bivariate distribution to the mean noise level of the combined channels, and the lower end by the combined noise.

The dynamic range of the combined channels is, surprisingly, the same as that for each separate channel. This is because although we have two noise contributions to the output, the maximum output pulse for a strong signal is associated with the corners of the output bivariate distribution where it is $\sqrt{2}$ times stronger than at the middles of the sides (see Fig. 9). Thus the output dynamic range is $\sqrt{n} : 1$ for both single and combined double channels. For a 32-bit code, for example, it is about 15 dB.

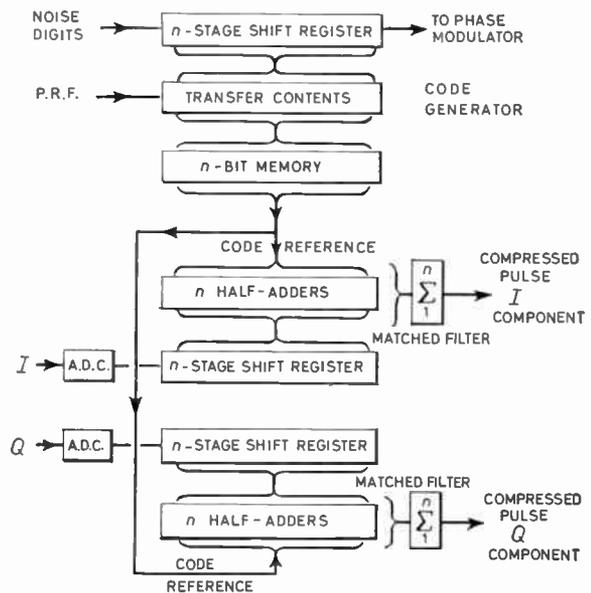


Fig. 11. Block diagram of a digital matched-filter system.

10.3. Overlapping Signals

The behaviour of the system under conditions of overlapping signals is complex, and depends on the degree of overlap, the relative sub-pulse timing and relative signal amplitude and phase-angles. In a simplified case we may have signals occupying the same quadrature channel, or different quadrature channels where there is a substantial 90° component in their relative phase angles. Some of the binary decisions in either or both channels may be reduced in accuracy so that the output bivariate binomial distribution is deflected not towards the corners (the most probable case when the same strong signal is identified in both quadrature channels) but towards somewhere along a side depending on the degree of overlap, and with a reduced magnitude of deflection. A deflection towards the middle of one side implies

a total loss of correctly identified digits in one channel, which will at some point furnish some other compressed pulse due to the overlapping signal.

In either channel, any correctly identified digits belonging to the signal will make an appropriate contribution to its own particular compressed pulse in one or both channels. The overlapping (random) sub-pulses of other signals will merely reduce the probability of making such correct identification, and are a favourable feature of randomly changing codes. Fixed coding would introduce the danger of fixed, i.e. repetitive side-lobe interaction between such overlapping signals. However, as has already been stated, the selection of transmitted codes need only be random over a period which exceeds the integration time employed in the receiver.

11. The Response to Doppler-shifted Signals

The use of random binary coding frees the designer from the limitation on code length hitherto imposed by code a.c.f. side-lobes. Instead, the maximum number of digits is set by the maximum clock-rate, and the duration of the long pulse. The latter is itself limited by the maximum required Doppler channel-width at the particular radar band. If the target velocity is too high, signal phase shift during the long pulse may exceed $\pm\pi/2$ radians, and the polarity of some digits may be inverted, so as to cause destructive cancellation instead of addition, reducing the response to sufficiently fast radially moving targets.

When the target velocity is such that $\pm\pi$ radians phase-shift occurs during the long pulse, half the resolved digits will cancel the other half, and a 'blind' velocity results. Further increase of target velocity (phase-shift $\pm 3(\pi/2)$ per long pulse) will produce a further, though lower peak of sensitivity, and so on. The use of binary, rather than n -ary phase-coding results in wider Doppler channels or less restricted long pulse duration.

The effect may be overcome (or perhaps even exploited?) by the provision of separate channels having either a frequency-offset detector system, or a Doppler-corrected pattern of connections of the digital matched filters. Each additional channel might be centred on the first null of its neighbour.

12. Conclusions

The technique described enables the matched-filter for digitally coded signals to be realized with physically small elements. The shift-register possesses the reliability of digital techniques, and obviates the complicated setting up procedures of analogue matched-filters.

High values of time-bandwidth product are readily obtainable and yet the range-sidelobe problems associated with long fixed codes are avoided.

Under certain conditions a linear relationship is found to exist between input and output signal to noise ratios. This is for the weak signal case, or when a long code, i.e. a large number of sub-pulses, shares the available signal energy. The maximum output signal/noise ratio and dynamic range are designable, being determined by the code length.

The system has a constant false alarm rate, determined by the code length and an output detection threshold.

Range indication is unambiguous. When there is no *a priori* knowledge of target range, an average loss of 3 dB occurs due to the mismatch between range resolution cells and the clock intervals. The system appears otherwise to be optimum for those targets whose geometry does not mutilate the code.

When signals are absent, the output of each digital matched-filter system has a random binomial distribution. It was the desire that the a.c.f. side-lobes of the transmitted waveform should resemble this 'no signal' condition in the neighbourhood of the compressed pulse which suggested the choice of random binary coding.

It has been possible, by digital computer simulation, to show that the selection of a group of random codes having 'good' side-lobes has little advantage to offer. The mutilation of the codes by noise-induced decision errors quickly destroys their favourable properties even at moderate signal/noise ratios, and randomly chosen codes (with no special selection of the better ones) will be equally effective in practice.

13. Acknowledgments

It is the author's pleasure to acknowledge the interest and encouragement of Mr. E. H. Boyenval and Mr. P. M. Woodward, and other colleagues at the Royal Radar Establishment.

Contributed by permission of the Director of the Royal Radar Establishment.

14. References

1. Matched Filter Issue, *Trans. Inst. Radio Engrs on Information Theory*, IT-6, No. 3, June 1960.
2. Skolnik, M. I., 'Introduction to Radar Systems', p. 416 (McGraw Hill, New York, 1962).
3. Meyer, P. A., 'Filters with an approximately rectangular impulse-response', *Nachrichtentechnische Zeitschrift*, 18, No. 5, p. 249, May 1965. Also *NTZ-CJ*, 1966, No. 3, p. 108 (English edn.).
4. Lawton, J. G., 'Comparison of binary data transmission system', Proceedings of 2nd I.R.E. National Convention on Military Electronics, p. 54, Appendix 1, 1958.

5. Woodward, P. M., 'Probability and Information Theory with Applications to Radar', pp. 6-7 (Pergamon Press, London, 1953).
6. Montgomery, G. F., 'A comparison of amplitude and angle modulation for narrow-band communications of binary coded messages in fluctuation noise', *Proc. I.R.E.*, pp. 447-454, eqn. (14), February 1954.
7. Davenport, W. B., Jr., 'Signal to noise ratios in band-pass limiters', *J. Appl. Phys.*, **24**, pp. 720-7, June 1953.
8. Wetherall, P. R., Whitfield, G. R. and Woodward, P. M., 'A comparison of two methods of detecting a signal of unknown phase in the presence of noise', *R.R.E. Journal*, No. 52, pp. 141-6, April 1965.
9. Gardner, F. M., 'Phaselock Techniques', p. 55 (Wiley, New York, 1966).

15. Appendix

Derivation of the expression for the output signal/noise ratio of a single-channel digital matched-filter when noise induced decision errors exist in the sequence being processed.

When no signal is present, $P_e = P_c = 0.5$
 variance of output binomial distribution
 $= nP_eP_c = n/4$
 standard deviation is then $(\sqrt{n})/2$ (voltage levels) ... (11)

mean of distribution is $n/2$
 When signal is present,
 signal output = deflected mean - undeflected mean
 $= nP_c - n/2$
 $= \frac{n}{2}(1 + \text{erf} \sqrt{(E/N_0)}) - n/2$
 $= \frac{n}{2} \text{erf} \sqrt{(E/N_0)}$ (voltage levels) ... (12)

where E/N_0 is the sub-pulse signal/noise energy ratio.
 Dividing eqn. (12) by eqn. (11).

output s.n.r. = $(\sqrt{n}) \text{erf} \sqrt{(E/N_0)}$ (13)

Manuscript first received by the Institution on 16th February 1968 and in final form on 6th May 1968.

(Paper No. 1196/AMMS13.)

© The Institution of Electronic and Radio Engineers, 1968

FORTHCOMING CONFERENCES

Gas-Filled Valves

A Conference on Gas-Filled Valves, to be held at the University of Southampton from 18th to 20th September 1968, is being organized by the Institute of Physics and The Physical Society. All types of gas-filled devices will be dealt with, covering both electronics and power applications and including thyratrons, mercury arc converters, ion engines and ion sources, together with fundamental aspects of electron emission conduction, recovery, gas/surface effects and beam/plasma interactions.

It is proposed to hold an exhibition of devices and experiments related to the Conference. A visit will be arranged to the Marchwood Engineering Laboratories of the C.E.G.B. to see work on high voltage mercury arc valves for d.c. power transmission.

Further details and application forms are available from the Meetings Officer, I.P.P.S., 47 Belgrave Square, London, S.W.1.

Education for Quality and Reliability

The 4th Residential Course of the National Council for Quality and Reliability will be held at Birmingham University from 4th to 16th August 1968. The course is designed for those responsible for planning educational activities in and for industry, in national and local government, and in educational establishments. It should also be of practical value to management. The inclusive fee is £100. Further information may be obtained from N.C.Q.R., Vintry House, Queen Street-place, London, E.C.4.

Underwater and Marine Technology

The Society for Underwater Technology is sponsoring the International Oceanology Conference at the International Oceanological Equipment and Services Exhibition at Brighton from 16th to 21st February, 1969. The conference is intended for industrialists, economists, scientists, engineers, and technologists from all nations interested in the exploitation of the oceans. The theme will be discussed under the following broad headings:

- Biological and mineral resources;
- Transport and transmission;
- Navigation, communication, meteorology;
- Coastal control and amenities;
- Economic prospects.

National contributions on industrial programmes, organizational problems and international collaboration have been arranged. Parallel technical sessions are also planned for which papers would be welcome on machines and techniques, particularly where these

show novel features or achievement of significantly lower cost of operation. Such papers will be grouped into sessions under the following headings:

Underwater (including control from surface) handling, working, measuring, viewing, sampling and sensing.

Underwater survey, positioning, navigation and communication.

Engineering for marine environment.

Data gathering, handling and use.

Instruments, sensors, systems and power supply.

The Conference Working Party which is under the Chairmanship of Professor D. G. Tucker (Fellow), invites offers of papers (of about 2,500 words) on suitable subjects, and Synopses of 150 words should be forwarded without delay. Completed texts are required by 1st October 1968.

The Oceanological Equipment Exhibition will cover all aspects of oceanological investigation and underwater technology, and will be accompanied by sea-borne demonstrations with the use of a variety of craft such as research vessels, hovercraft, hydrofoils, helicopters and amphibious vehicles.

Further information is available from the Conference Secretariat, Society for Underwater Technology, BPS Exhibitions Ltd., 6 London Street, London, W.2.

Microelectronics Conference

The Third International Conference on Microelectronics will be held in conjunction with the Electronica Trade Fair in Munich, from 11th to 13th November 1968. The two previous conferences, in 1964 and 1966, attracted a large attendance, and it is expected that approximately 1000 delegates will participate in the 1968 Conference. The following topics will be dealt with:

Materials and procedures: new semiconducting materials, organic semiconductors, bulk-effect devices; new procedures for manufacturing integrated circuits, application to mass production, computer-aided design, packaging of the first and second level.

Circuits and applications: new digital and linear circuits, functional i.c.s, systems approach and problems with l.s.i., monolithic memories.

Impact of microelectronics on conventional components: hybrid circuits, packaging.

Optoelectronics: light modulation, display, memories.

Further information on the Conference and details of registration may be obtained from the Congress Bureau, Theresienhöhe 15, 8000 Munich 12, Germany.

The Design of Low-noise Audio-frequency Amplifiers

By

E. A. FAULKNER,
M.A., Ph.D., C.Eng., F.I.E.R.E.,
M.I.E.E.†

Summary: This paper contains: (a) an account of the phenomenological theory of noise in linear amplifiers operating from resistive signal sources; (b) a discussion of the noise parameters of bipolar and junction field-effect transistors; and (c) examples of how these principles can be applied in practical circuit design for resistive sources.

List of Contents

1. Introduction
2. Noise in Linear Amplifiers
 - 2.1. Signal/noise ratio and noise figure
 - 2.2. Models of noisy amplifiers
 - 2.3. Variation of F with R_s . Equivalent noise voltage and current generators
 - 2.4. Noise resistance, optimum source resistance and minimum noise figure
 - 2.5. Noise-matching
 - 2.6. Noise figure of cascaded amplifiers
 - 2.7. Measurement of noise parameters
3. Transistor Noise
 - 3.1. Noise parameters of bipolar transistors
 - 3.2. Noise figure and noise resistances of bipolar transistors
 - 3.3. Noise parameters of junction f.e.t.s
 - 3.4. Comparison of j.f.e.t.s and bipolar transistors
4. Negative Feedback and Noise Figure
 - 4.1. Calculation of noise figure
 - 4.2. Practical application of the theory
 - 4.3. CE, CC and CB connections
 - 4.4. Use of negative feedback to improve noise figure
5. Practical Circuit Design
 - 5.1. Worst-case design of the input stage
 - 5.2. The feedback resistor
 - 5.3. Worst-case design of the second stage
 - 5.4. General comments on circuit design
6. Acknowledgment
7. References

1. Introduction

There are numerous published accounts of the theory of noise in transistor amplifiers, but it is evident that many circuit designers are not fully aware of the engineering realities which underlie the algebraic formalities. In this work we shall attempt

† J. J. Thomson Laboratory, University of Reading, Whiteknights, Reading, Berkshire.

to show the way in which the theory may be applied to the practical design of low-noise amplifiers for the audio-frequency range. To begin, we shall develop the basic theory in a way which is intended to emphasize the physical principles.

Throughout the discussion, we shall be assuming that the impedance of the signal source is resistive.

2. Noise in Linear Amplifiers

2.1. Signal/Noise Ratio and Noise Figure

In Fig. 1(a) is shown a signal source, whose impedance will be assumed to be passive, resistive and equal to R_s , connected to a noisy infinite-impedance linear voltage amplifier of voltage gain A . The equivalent circuit shown in Fig. 1(b) represents the source as two voltage generators in series with a noiseless resistance. The first of these voltage generators, v_{NR} , is the Johnson noise voltage generated in the resistance R_s , and has therefore a mean-square value of $4R_s kT\Delta f$, where Δf is the frequency range being considered expressed in Hz; k is Boltzmann's constant in joule/degK, and T is the absolute temperature in degrees K.‡ The second generator v_s is

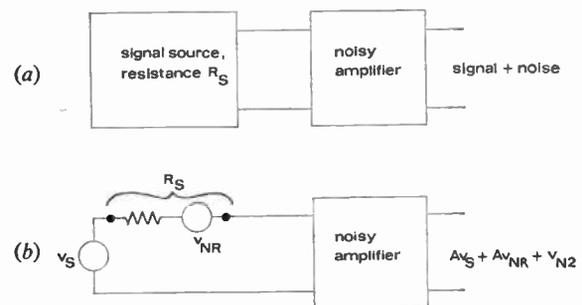


Fig. 1. Noise figure of an amplifier.

$$F = 1 + \frac{v_{N2}^2}{4A^2 R_s kT \Delta f}$$

‡ For $R_s = 1 \text{ k}\Omega$ and $\Delta f = 1 \text{ kHz}$, this gives $0.13 \text{ }\mu\text{V}$ r.m.s. at 300°K . Note that in the case of a broad-band amplifier with 6 dB/octave roll-off, the equivalent noise bandwidth is greater than the 3 dB bandwidth by a factor $\pi/2$.

called the *signal* and includes all the remainder of the voltage generated in the source; for the purpose of this discussion it is simplest for us to visualize the signal as a purely sinusoidal generator, but in fact it may contain random components (for instance, from a noisy transducer or transmission line) and may even be completely random (for instance, where the amplifying system is being used to study a random process). The output of the amplifier contains three distinct components: an amplified signal voltage Av_S , an amplified noise voltage Av_{NR} , and an additional random voltage v_{N2} which is the noise contributed by the amplifier itself. There will be no correlation between v_{N2} and v_{NR} , so the total mean-square noise voltage at the output is $\overline{v_{N2}^2} + A^2\overline{v_{NR}^2}$. The noise figure of the system is then given by

$$\begin{aligned}
 F &= \frac{\text{best possible signal/noise ratio}}{\text{actual output signal/noise ratio}} \\
 &= \frac{\overline{v_S^2}/\overline{v_{NR}^2}}{A^2\overline{v_S^2}/(A^2\overline{v_{NR}^2} + \overline{v_{N2}^2})} \\
 &= 1 + \frac{\overline{v_{N2}^2}}{4A^2R_S kT\Delta f} \dots\dots(1)
 \end{aligned}$$

Equation (1) may be taken as a general definition of F .† It is important to notice that the noise generated by the amplifier can be formally represented as being due to a generator v_{N2}/A connected in series with the signal input.

As a practical guide, we may regard any system having a noise figure of 3 dB or better ($F \leq 2$) as being a *low-noise* system.

2.2. Models of Noisy Amplifiers

It will help us to understand the properties of actual amplifiers if we first set up some models which consist of idealized noiseless amplifiers in conjunction with resistances and/or generators connected to the input.

In Figs. 2(a), 2(b), 2(c), the amplifier is assumed to have infinite input impedance and to be noiseless so that the signal/noise ratio at the output terminals is the same as that at the input terminals.

In Fig. 2(a) a signal source, represented by a generator v_S in series with a resistance R_S , is connected directly to the amplifier, the input signal/noise ratio is $\overline{v_S^2}/4R_S kT\Delta f$ and the noise figure is of course unity. In Fig. 2(b) a resistor R_1 has been connected in series with the amplifier. Because of the infinite input

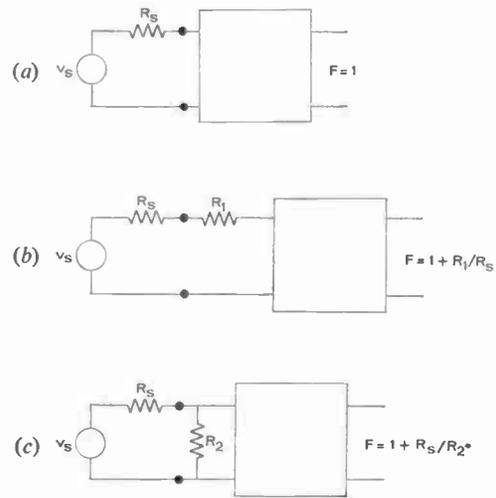


Fig. 2. Effect of series and parallel resistance on noise figure. The 'black box' represents a noiseless amplifier.

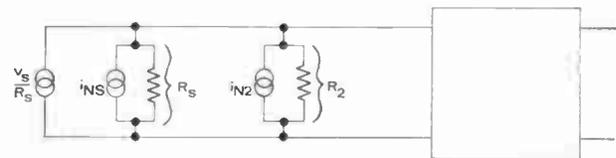


Fig. 3. Norton equivalent circuit of Fig. 2(c).

$$F = (\overline{i_{NS}^2} + \overline{i_{N2}^2})/\overline{i_{NS}^2} = 1 + (R_S/R_2).$$

impedance of the amplifier, R_1 has no effect on the input signal voltage but the mean-square input noise voltage increases to $4(R_S + R_1)kT\Delta f$; by the method used in equation (1) we find that

$$F = 1 + (R_1/R_S) \dots\dots(2)$$

Figure 2(c) shows a resistor R_2 connected in parallel with the amplifier input terminals. This has the effect of reducing the noise voltage at the amplifier input, since the effective resistance is now the parallel combination of R_S and R_2 , and its mean-square value is now $4kT\Delta f R_S R_2 / (R_S + R_2)$. However, the mean-square signal voltage at the amplifier input is reduced by the factor $R_2^2 / (R_S + R_2)^2$, and the overall result is a deterioration in the signal/noise ratio by the factor $R_2 / (R_S + R_2)$. The corresponding noise figure is given by

$$F = 1 + (R_S/R_2) \dots\dots(3)$$

This result can be obtained more conveniently by the use of the Norton equivalent circuit shown in Fig. 3. The signal, the noise in R_S , and the noise in R_2 are represented by parallel current generators and we arrive at equation (3) for the noise figure by

† There are some who hold that F should be called 'noise figure' only when it is expressed in dB, and should otherwise be referred to as 'noise factor'. Such a distinction seems impossible to justify on a logical basis.

considering current ratios, without explicitly referring to the magnitudes of the voltages appearing at the amplifier input terminals.

A great deal of confusion has arisen over the relation between noise figure and input resistance. It should be clear from Fig. 3 that if an amplifier has a parallel input resistor R_2 , the effect of this on the noise figure can be completely described in terms of the noise current which it generates, without consideration of the simple shunting effect of the resistor on the actual magnitude of the signal or the noise. No passive resistor R can develop a mean-square noise current less than $4kT\Delta f/R$ (it may develop more noise if it has direct current flowing through it) and accordingly it is not possible for an amplifier to give a noise figure close to unity unless its passive input resistance is substantially greater than the source resistance. On the other hand, an additional noiseless resistance connected across the amplifier terminals in Fig. 3 will have no effect on the signal/noise ratio. We see therefore that when the input resistance is determined by the action of active components (for instance, by means of parallel feedback) it is possible in principle to make it much less than the source resistance without degrading the noise figure.

A point which may seem surprising in this discussion is the apparently quite dissimilar roles played by the series resistor R_1 in Fig. 2(b), which increases the noise without affecting the signal, and the resistor R_2 in Fig. 2(c), which reduces the noise but reduces the signal still more. The reason for this asymmetry is our choice of the amplifier as an ideal voltage amplifier of infinite input resistance. If we had chosen an ideal current amplifier with zero input resistance (infinite input conductance) then R_1 rather than R_2 would have been the one which affected the signal magnitude; expressions (2) and (3) for the noise figures would of course be unchanged. If we take the intermediate case and assume the amplifier to have a finite, though still noiseless, input resistance, then both R_1 and R_2 will have an effect on the signal magnitude, the noise figures still being unchanged.

2.3. Variation of F with R_s . Equivalent Noise Voltage and Current Generators

In Section 2.2 we have considered two simple models of noisy amplifiers. A noiseless amplifier in series with a resistor R_1 becomes a noisy amplifier with noise figure $[1 + (R_1/R_s)]$, and a noiseless amplifier in parallel with a resistor R_2 becomes a noisy amplifier with noise figure $[1 + (R_s/R_2)]$. Now if we calculate the way in which the noise figure of the most general linear amplifier can vary with R_s , making no assumptions apart from that of linearity, we find the relation

$$F = 1 + k_1/R_s + R_s/k_2 + k_3 \dots(4)$$

where k_1 , k_2 and k_3 are constants of the amplifier, all being in general functions of frequency. Equations (2) and (3) are, of course, special cases of equation (4).

Now let us consider the equivalent circuit shown in Fig. 4(a), in which a noisy amplifier is represented as a noiseless amplifier with a random voltage generator v_{NA} , and a random current generator i_{NA} , connected to its input terminals. Figure 4(b) shows the same circuit with all the noise sources, including the Johnson noise in R_s , shown as voltage generators in series with the input. In calculating the resultant noise voltage we must bear in mind that when two random noise generators v_{N1} and v_{N2} are connected in series, the resulting voltage v_N has a mean-square value given by

$$\overline{v_N^2} = \overline{v_{N1}^2} + \overline{v_{N2}^2} + 2\gamma(\overline{v_{N1}v_{N2}})^{\frac{1}{2}} \dots(5)$$

where γ is a parameter called the correlation coefficient between the two generators, having some value between -1 and $+1$. Thus we obtain for the noise figure F :

$$F = 1 + \frac{\overline{v_{NA}^2}}{4R_s k T \Delta f} + \frac{R_s \overline{i_{NA}^2}}{4k T \Delta f} + 2\gamma \frac{(\overline{v_{NA}^2} \overline{i_{NA}^2})^{\frac{1}{2}}}{4k T \Delta f} \dots(6)$$

This equation has the same form as equation (4), having one term in R_s , one term in $1/R_s$ and one term independent of R_s . It follows that the equivalent circuit shown in Fig. 4(a) is an appropriate way of representing equation (4), and that the parameters k_1 , k_2 and k_3 can be specified by specifying $\overline{v_{NA}^2}$, $\overline{i_{NA}^2}$ and γ . The result given in equation (6) is quite independent of the input resistance of the amplifier,

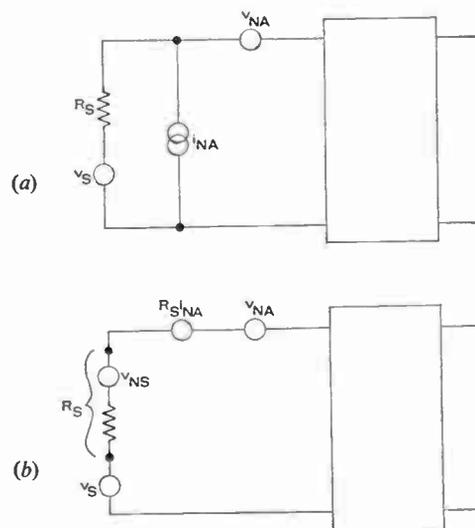


Fig. 4. Equivalent circuits of noisy amplifier. The 'black box' represents a noiseless amplifier.

provided that it can be assumed noiseless—that is, provided that the noise generators v_{NA} and i_{NA} include *all* the noise sources in the amplifier. It can be shown from linear-circuit theory that the value of k_3 is never great enough to require the magnitude of γ to be greater than unity.

It is important to notice that the equivalent generators in Fig. 4(a) are a purely formal way of representing the variation of F with R_S for a given amplifier, and include the effects of all the noise sources in the amplifier whether or not they originate in the input circuit. We can, for instance, choose values of $\overline{v_{NA}^2}$, $\overline{i_{NA}^2}$ and γ such that the mean-square noise voltage at the amplifier terminals is quite independent of the value of R_S . If we assume the input impedance of the amplifier to be a resistance r_i , we find that the required relation is

$$\overline{v_{NA}^2}/\overline{i_{NA}^2} = r_i^2, \quad \gamma = 1 \quad \dots\dots(7)$$

These conditions are a means of representing an amplifier in which the noise comes entirely from the output stages.

2.4. *Noise Resistance, Optimum Source Resistance and Minimum Noise Figure*

The equivalent input noise voltage and current generators are nowadays widely used for specifying the noise performance both of complete amplifiers and of input devices. The units commonly used are the ‘nanovolt per square-root-hertz’ and the ‘picoampere per square-root-hertz’ which are somewhat cumbersome and uninformative. For most applications, and certainly in the field of a.f. amplifier design, it is much more satisfactory to use the *series noise resistance* R_{Nv} and the *parallel noise resistance* R_{Ni} . These quantities are defined by the relations

$$R_{Nv} = \overline{v_{NA}^2}/4kT\Delta f, \quad R_{Ni} = 4kT\Delta f/\overline{i_{NA}^2} \quad \dots\dots(8)$$

and it is convenient to remember that a noise resistance of $x \text{ k}\Omega$ is equivalent to a voltage generator of $4\sqrt{x}$ nanovolt, or a current generator of $4/\sqrt{x}$ picoampere, per square-root-hertz at 300°K .

Using the noise resistances we can rewrite equation (6) in the more convenient form

$$F = 1 + R_{Nv}/R_S + R_S/R_{Ni} + 2\gamma\sqrt{R_{Nv}/R_{Ni}} \quad \dots\dots(9)$$

This expression is obviously related to equations (2) and (3), which refer to the simplest models of noisy amplifiers.

The idea of an ‘equivalent noise resistance’ has long been used in connection with vacuum tubes, but this concept has traditionally referred to an equivalent series noise resistance only, and it follows from (9) that if we restrict ourselves to one equivalent noise resistance rather than two, we must make it

depend on R_S . We shall see later that the term in γ can in many practical cases be ignored, so that the noise performance of a practical low-noise amplifier, or input transistor, will normally be specified by two resistances. These equivalent noise resistances will, of course, be functions of the frequency.

The formulation of equation (9) makes it particularly easy to understand how F varies with R_S . Because the equation contains terms in R_S and in $1/R_S$, F becomes very large in the limits of large and small R_S and has a minimum for an optimum value of R_S . Differentiation with respect to R_S shows that the minimum occurs when the two terms containing R_S are equal, and we obtain for the optimum source resistance

$$(R_S)_{opt} = \sqrt{R_{Nv}R_{Ni}} \quad \dots\dots(10)$$

the optimum being the geometric mean of the two equivalent noise resistances. By substituting this value in equation (9) we may calculate the minimum noise figure, that is the value of F corresponding to $(R_S)_{opt}$:

$$F_{min} = 1 + 2(1 + \gamma)\sqrt{R_{Nv}/R_{Ni}} \quad \dots\dots(11)$$

We see that a good noise figure is obtainable if the series noise resistance is small compared with the parallel noise resistance, so that a value of R_S can be chosen which satisfies the condition

$$R_{Ni} \gg R_S \gg R_{Nv}$$

and the noise resistances have only a small effect in the input circuit.

In practical a.f. work, it is usually true to say that a noise figure of 1 dB = 1.26 is indistinguishable from the ‘best possible’ figure of 0 dB = 1.0, and equation (11) shows that this figure can be achieved for $R_{Ni} = 60 R_{Nv}$ if γ is assumed to be zero. As we shall see later, this condition is easily satisfied by a high-gain bipolar transistor or a junction field-effect transistor, under the best operating conditions and in the best part of its frequency range. In fact, the assumption $\gamma = 0$ is correct in these cases, so that we can completely specify the noise performance at a given frequency by specifying the minimum noise figure F_{min} and the optimum source resistance $(R_S)_{opt}$. In terms of these parameters we can express equation (9) in the form

$$(F - 1) = \frac{1}{2}(F_{min} - 1) \left[\frac{(R_S)_{opt}}{R_S} + \frac{R_S}{(R_S)_{opt}} \right] \quad \dots\dots(12)$$

It should be clear from this discussion that the low-noise capability of an amplifier, or of a transistor, in a given frequency range should be assessed on the basis of its minimum noise figure, rather than the noise figure obtained from an arbitrarily chosen source resistance, or an apparently ‘low’ value of series noise resistance.

Although a noise figure of 1 dB may be practically indistinguishable from 0 dB, it is sometimes very advantageous to use an input device with a minimum noise figure of much less than 1 dB, because such a device will maintain a satisfactory noise figure over a comparatively wide range of source resistance.

2.5. Noise-matching

Suppose that we have available a low-noise amplifier whose optimum source resistance $(R_s)_{opt}$ differs widely from the actual resistance R_s of the signal source to be used. In principle we can always 'noise-match' the amplifier to the source by using an ideal input transformer of ratio n ; this reflects into the secondary circuit a signal voltage equal to nv_s , where v_s is the input signal voltage generator, and a resistance equal to n^2R_s . The 'best possible' signal/noise ratio then remains unchanged at $\overline{v_s^2}/4R_s kT\Delta f$, and accordingly we can improve the overall noise figure by choosing the transformer ratio so that n^2R_s approximates to $(R_s)_{opt}$.

In practice, it is often preferable to avoid the use of an input transformer for noise-matching in audio-frequency circuits unless a transformer is essential for some other reason such as d.c. isolation. In cases where $R_s < (R_s)_{opt}$ one can in principle achieve the same noise-figure improvement as is obtainable from an input transformer of ratio n (n being an integer) by the technique of using n^2 identical amplifiers connected in parallel; the combined amplifier has series noise resistance and parallel noise resistance both reduced by a factor n^2 compared with an individual amplifier.¹ The minimum noise figure is therefore unchanged, but the optimum source resistance reduced by a factor n^2 .

2.6. Noise Figure of Cascaded Amplifiers

An important situation is that in which two amplifiers are connected in cascade, that is with the output of the first acting as the input of the second.

Suppose that the first amplifier is driven by a signal source of resistance R_1 , and has a noise figure F_1 and voltage gain A_1 referred to this source resistance. The mean-square value of the noise voltage v_{N1} generated in the output circuit of the first amplifier is then given by

$$\overline{v_{N1}^2} = 4R_1 kT\Delta f \cdot F_1 A_1^2 \quad \dots\dots(13)$$

Now the noise voltage v_{N2} appearing in the output circuit of the second amplifier is the sum of two components: (i) a term equivalent to (13) containing the noise figure F_2 and voltage gain A_2 of the second amplifier referred to its source resistance R_2 , which is in fact that output resistance of the first amplifier;

and (ii) the noise voltage resulting from the amplification of v_{N1} :

$$\overline{v_{N2}^2} = 4kT\Delta f A_2^2 (R_2 F_2 + R_1 F_1 A_1^2) \quad \dots\dots(14)$$

We thus obtain for the overall noise figure F the expression

$$F = F_1 + R_2 F_2 / R_1 A_1^2 \quad \dots\dots(15)$$

This result shows that, provided the ratio $A_1^2 R_1 / R_2$ is sufficiently large, the noise figure of the combined system becomes substantially equal to the noise figure of the first amplifier. This is of practical importance when $F_2 > F_1$, and is the principle of the *low-noise preamplifier*.

A convenient way of regarding the action of a preamplifier is to consider the equivalent input noise generators of the second amplifier, expressed as a single input voltage generator v_{NA2} in relation to its source resistance R_2 , transferred back to the input circuit of the first amplifier. The resulting generator is v_{NA2}/A_1 , and from this point of view we see that the function of the preamplifier is to reduce the equivalent effect on the input circuit of the noise generated by the second amplifier.

2.7. Measurement of Noise Parameters

If we wish to measure F for a given amplifying system, it may be necessary first to limit the frequency response to the required range by means of filters. A small signal v_s at the mid-band frequency is then introduced from a signal source of the required resistance R_s , and the output signal/noise ratio is measured and compared with the calculated 'best possible' value $\overline{v_s^2}/4R_s kT\Delta f$.

It is important to notice that a true-mean-square or true-r.m.s. measuring system must be used to obtain the signal/noise ratio. The use of standard noise sources, which in principle avoids the necessity for a measuring system of this type (and also avoids the necessity for an accurate knowledge of the bandwidth), is not usually desirable in the audio-frequency range; this is partly because of their limited accuracy, and partly because there is in principle no reason to suppose that their output has the same spectral distribution as has the noise to be measured.

In order to measure the equivalent noise generators and their correlation coefficient we must make three separate measurements of F at three different values of R_s . A measurement with a very low value of R_s gives R_{Nv} directly, the second term in (9) being the dominant one. Similarly a measurement with a very high value of R_s gives R_{Ni} directly. The correlation coefficient γ may conveniently be evaluated from a measurement of F with R_s approximately equal to its optimum value $\sqrt{(R_{Nv} R_{Ni})}$, but in most practical

cases γ is known to be zero. When performing these measurements one must bear in mind that the bandwidth may be strongly dependent on the source resistance.

3. Transistor Noise

3.1. Noise Parameters of Bipolar Transistors

The noise sources, apart from flicker noise, in bipolar transistors have been discussed by van der Ziel.^{2,3} His conclusions can be expressed approximately in the form of a simplified equivalent circuit.

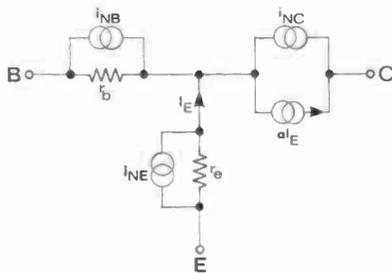


Fig. 5. Simplified equivalent circuit of bipolar transistor including noise generators.

In the circuit shown in Fig. 5, the transistor action is represented by the resistors r_b and r_e and the current generator αI_E . The transistor capacitances, and also the collector resistance, have been omitted. Three noise-current generators are shown, with mean-square values given by:

$$\left. \begin{aligned} \overline{i_{NE}^2} &= 2qI_E \Delta f \\ \overline{i_{NB}^2} &= 4kT \Delta f / r_b \end{aligned} \right\} \dots\dots(16)$$

$$\overline{i_{NC}^2} = 2qI_C [1 - |\alpha|^2 / \alpha_0] \Delta f \dots\dots(17)$$

For our present purposes, any correlation between these generators can be ignored.

The generator i_{NE} can be called 'shot noise' because its value is given by the same formula as for a temperature-limited vacuum diode. The generator i_{NB} represents the Johnson noise in the base resistance r_b . The generator i_{NC} was labelled 'partition noise' by van der Ziel because in the low-frequency limit it follows the same equation as does partition noise in a vacuum tube; but its mean-square value increases with frequency as $|\alpha|$ decreases and it is responsible for the falling-off in the noise performance of the transistor at high frequencies. A simple calculation shows that the i_{NC} generator is 3 dB above its low-frequency value at an angular frequency of $\omega_T / \sqrt{\beta_0}$, and then increases at 6 dB per octave. We notice that the 'corner frequency' for this noise generator is approximately the geometric mean of the common-base and the common-emitter cut-off frequencies,

the latter quantity being ω_T / h_{fe0} and h_{fe0} being approximately equal to $\beta_0 \cdot \dagger$

In practical audio-frequency designs we find that if we use modern silicon planar transistors, which are the only type of small-signal bipolar transistor that interests us here, we can usually ignore the frequency variation of i_{NC} expressed in equation (17), because $|\alpha|$ is always substantially equal to α_0 . Experimentally we find that Fig. 5, in conjunction with equations (16) and (17), gives a satisfactory description of the noise behaviour of the transistor at the upper end of the audio-frequency range, but $\overline{i_{NC}^2}$ increases at low frequencies. The additional contribution to the noise at low frequencies is variously described as 'flicker noise', 'excess noise', or '1/f noise' and we can express this effect to a reasonable approximation by means of the equation

$$\begin{aligned} \overline{i_{NC}^2} &= 2qI_C (1 - \alpha_0) (1 + \omega_F / \omega) \Delta f \\ &= 2q\alpha_0 I_C (1 + \omega_F / \omega) \Delta f / \beta_0 \dots\dots(18) \end{aligned}$$

where ω_F is a parameter which may be called the flicker-noise characteristic frequency.

A noise source which has not yet been mentioned is the noise generated by the collector leakage current.³ For modern small-signal silicon planar transistors operated at d.c. collector currents of 100 nA and above, this noise source is negligible; we shall therefore not discuss it further here.

3.2. Noise Figure and Noise Resistances of Bipolar Transistors

From the equivalent circuit of Fig. 5 in conjunction with equations (16) and (18), we can derive the following expression for the noise figure F in the common-emitter configuration:

$$F = 1 + \frac{r_b + r_e / 2}{R_S} + \frac{(R_S + r_b + r_e)^2 (1 + \omega_F / \omega)}{2\beta_0 r_e R_S} \dots\dots(19)$$

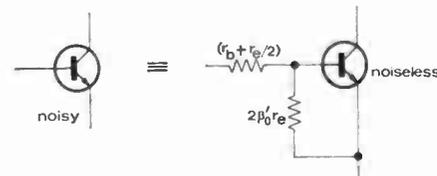


Fig. 6. Equivalent noise resistances of a bipolar transistor, assuming conditions (21) to be satisfied.

This formula is applicable in the frequency range well below $\omega_T / \sqrt{\beta_0}$.

[†] We use the symbol β_0 for the direct current gain, and the symbol h_{fe0} for the low-frequency value of the alternating current gain. We use the term 'frequency' to refer either to the cyclic frequency f or the angular frequency ω , according to the context.

In order to interpret equation (19) in relation to equivalent series and parallel noise resistances, we must separate it into the terms in R_s , the terms in $1/R_s$, and the terms independent of R_s . We see that the effect of the flicker-noise factor in (19) is formally equivalent to a decrease in β_o as the frequency falls below ω_F , and we shall find it convenient to make use of the parameter β'_o defined by the relation

$$\beta'_o = \beta_o / (1 + \omega_F / \omega) \quad \dots\dots(20)$$

If now we can make the assumptions

$$\beta'_o \gg 1 \text{ and } r_e \gg r_b \quad \dots\dots(21)$$

we obtain from (19) the expression

$$F \simeq 1 + \frac{r_b + r_e/2}{R_s} + \frac{R_s}{2\beta'_o r_e} \quad \dots\dots(22)$$

Equation (22) is a very important one. It states that, provided conditions (21) are satisfied, the noise generators in a bipolar transistor are equivalent to two uncorrelated generators which can be specified as a series noise resistance R_{Nv} and a parallel noise resistance R_{Ni} :

$$\begin{aligned} R_{Nv} &= r_b + r_e/2 \\ R_{Ni} &= 2\beta'_o r_e \\ \gamma &= 0 \end{aligned} \quad \dots\dots(23)$$

This result is illustrated in Fig. 6.

We can more easily appreciate the meaning of equations (23) and Fig. 6 if we take a concrete example. Suppose that we have a transistor with $\beta_o = 100$ independent of I_C , and r_b equal to 200Ω —a typical value for a modern transistor designed for low-noise a.f. applications. We remind ourselves that r_e is about 25Ω at room temperature for I_C equal to 1 mA , and that r_e varies inversely with I_C . If we are considering the frequency range above the flicker-noise region, conditions (21) are satisfied for values of I_C up to about $100 \mu\text{A}$. At this operating current the series noise resistance is 325Ω , and the parallel noise resistance is $50 \text{ k}\Omega$; from (10) and (11) we see that the optimum source resistance is about $4 \text{ k}\Omega$, and the corresponding minimum noise figure is about 0.65 dB . Let us now put I_C equal to $1 \mu\text{A}$; r_e becomes $25 \text{ k}\Omega$ and the series noise resistance about $12.5 \text{ k}\Omega$, the contribution of r_b being negligible. The parallel noise resistance is $5 \text{ M}\Omega$, the optimum source resistance $250 \text{ k}\Omega$, and the minimum noise figure 0.4 dB . We can easily see from equations (23), (10) and (11) that when r_e is high enough (that is, when I_C is low enough) to make the effect of r_b negligible, the optimum source resistance is $r_e \sqrt{\beta_o}$, and the minimum noise figure is $(1 + 1/\sqrt{\beta_o})$.

It is interesting to notice that the low-frequency common-emitter input resistance of the transistor, which is given by $(r_b + h_{fe0} r_e)$, is approximately equal to the parallel noise resistance because, generally speaking,⁴ h_{fe0} lies somewhere between β_o and $2\beta_o$.

It follows that a good noise figure is not obtainable from a bipolar transistor in the common-emitter configuration unless it is operated under voltage-amplifier conditions—that is, with its input resistance substantially greater than the source resistance. A great deal of misunderstanding has arisen from the assertion made by nearly every author of texts on transistor circuit design that the bipolar transistor is ‘basically a current amplifier.’

3.3 Noise Parameters of Junction F.E.T.s

In audio-frequency amplifiers, the junction f.e.t. sometimes provides a satisfactory alternative to the bipolar transistor as an input device. The noise mechanism in these devices has been discussed by van der Ziel,⁵ whose conclusions can be summarized by the statement that in the absence of flicker noise the series and parallel noise resistances are given by the equations

$$\begin{aligned} R_{Nv} &= 0.7/g_m \\ R_{Ni} &= 2kT/qI_G \\ \gamma &= 0 \end{aligned} \quad \dots\dots(24)$$

where g_m is the low-frequency value of the transfer admittance, and I_G is the reverse gate current under the specified operating conditions. We have not been able to confirm these conclusions experimentally with the same degree of precision as that which applies to equations (23) for the bipolar transistor. However, it can easily be confirmed that the expression for R_{Nv} in equations (24) gives the correct order of magnitude, and that it correctly shows the general principle that the lowest value of R_{Nv} is obtained by operating the device with the highest possible value of g_m —in practice, this means with the highest possible value of drain current. For low-cost j.f.e.t.s at the present time, typical operating values of g_m are in the region $1\text{--}5 \text{ m}\Omega^{-1}$, and R_{Nv} at frequencies above the flicker-noise region is in the region $200 \Omega\text{--}1 \text{ k}\Omega$.

The expression for R_{Ni} in (24) predicts low-frequency values in excess of $50 \text{ M}\Omega$ for typical j.f.e.t.s, and this general conclusion can easily be confirmed experimentally; but for various practical reasons it is difficult to measure the exact dependence of R_{Ni} on I_G .

The effect of flicker noise in j.f.e.t.s is quite different from that in bipolar transistors in that it is R_{Nv} rather than R_{Ni} which deteriorates at the lower end of the frequency scale. To include this effect in equations (24) we may rewrite the first of these equations in the form

$$R_{Nv} = 0.7(1 + \omega_F/\omega)/g_m \quad \dots\dots(25)$$

although this form of frequency dependence is only an approximate representation of what is found in practice. The actual way in which R_{Nv} increases as ω is reduced may vary considerably, even between individual specimens of the same type of j.f.e.t., and also

depends on the value of the drain current; in the flicker-noise region there is likely to be an optimum operating current for each individual device, which varies according to the frequency range being considered.⁶

By paying sufficient money (up to £5 at the time of writing, February 1968) one can obtain selected j.f.e.t.s with values of R_{Nv} less than 25 kΩ at 10 Hz, and Knott⁷ has reported one specimen with R_{Nv} equal to 1 kΩ at 10 Hz and 300 Ω at 1 kHz. However, many low-cost j.f.e.t.s have a series noise resistance in the region of 1 MΩ at 10 Hz.

3.4. Comparison of J.F.E.T.s and Bipolar Transistors

It has already been pointed out that the low-noise capabilities of a device should, strictly speaking, be assessed on the basis of the minimum noise figure, which depends on the ratio of R_{Nv} and R_{Ni} according to equation (11), which may be expressed in the form

$$F_{min} = 1 + 2\sqrt{(R_{Nv}/R_{Ni})}$$

if the correlation coefficient γ is assumed to be zero. This condition is in fact always true for a practical j.f.e.t., and also for a practical bipolar transistor which is being operated under low-noise conditions.

Now for a bipolar transistor operated at a sufficiently low value of collector current to ensure that the effect of the base resistance is negligible, we see from equations (23) that the expression for F_{min} reduces to $(1 + 1/\sqrt{\beta'_0})$ if $\beta'_0 \gg 1$. It is currently possible to obtain devices with β'_0 in the region of 400 under these conditions throughout the audio-frequency range, and the corresponding value of F_{min} is $1.05 = 0.2$ dB. To estimate the corresponding figure for a j.f.e.t., we may assume the correctness of equations (24) and substitute the somewhat favourable values of $g_m = 5 \text{ m}\Omega^{-1}$ and $I_G = 10^{-10}$ A. We then obtain $R_{Nv} \simeq 200 \text{ }\Omega$ and $R_{Ni} = 500 \text{ M}\Omega$. With these assumptions the value of F_{min} is $1.0013 = 0.006$ dB. Even if we had assumed an unfavourable value of R_{Nv} , taking for example the 1 MΩ at 10 Hz mentioned above, we should have obtained a calculated value of $1.1 = 0.4$ dB.

It is clear from these considerations that the j.f.e.t. is inherently a much lower-noise device than the bipolar transistor. However, in audio-frequency applications with a resistive signal source the advantage of having a better value of F_{min} is usually an illusory one, because with typical devices it is only obtained in conjunction with source resistances which are greater than those normally encountered in practice, and which cannot be achieved by the use of a transformer because of capacitive effects. Even when one has a source of very high resistance, one must take into account the input capacitance of the device itself, which was not included in the expression for R_{Ni} given in equations (24). These facts must be

considered in relation to the fact that, in most practical audio-frequency applications involving a resistive source, a noise figure of 1 dB or less is indistinguishable from 0 dB. However, where the source is a capacitive one the f.e.t. may show considerable advantages over the bipolar transistor. Also, in instruments which are required to give a good noise figure over a wide range of source resistance the f.e.t. may be the best choice of input device, and for these applications the circuit is sometimes noise-matched to an acceptably low value of source resistance by the parallel-input technique discussed in Section 2.5.

4. Negative Feedback and Noise Figure

4.1. Calculation of Noise Figure

In calculating the overall noise figure of a negative-feedback system, the important first step is to represent *all* the noise sources, and also the operation of the feedback, as equivalent generators in the input circuit. If the amplifier employs series feedback, we use the voltage-generator representation shown in Fig. 7(a). The amplifier noise is represented as a single generator v_{N1} ; in this case it is not necessary to use two generators to represent the noise, because we are assuming the value of R_s to be fixed. The generators v_s and v_{NS} represent the signal and the source noise respectively, and the generator v_F represents the feedback.

Now if A is the voltage gain of the amplifier referred to the input terminals, and β is the feedback ratio, we have

$$v_F = A\beta v_1 \quad \dots\dots(26)$$

where v_1 is the voltage across the equivalent input terminals. Assuming that the equivalent amplifier has a noiseless input impedance Z_i (this implies that the generator v_{N1} includes the noise in the input resistor) we may write

$$v_1 = (v_s + v_{NS} + v_{N1} + v_F)Z_i / (R_s + Z_i) \quad \dots\dots(27)$$

and combining this with equation (26) we obtain

$$v_1 = (v_s + v_{NS} + v_{N1})Z_i / (R_s + Z_i)(1 - A'\beta) \quad \dots\dots(28)$$

where $A' = AZ_i / (R_s + Z_i)$

Equation (28) will give us the signal/noise ratio at the input terminals of the equivalent noiseless amplifier and hence the output signal/noise ratio. Before discussing it in more detail, we shall set up the corresponding equation for a parallel-feedback amplifier.

In Fig. 7(b) the signal, the source noise, the amplifier noise and the feedback are shown as current generators i_s , i_{NS} , i_{N1} , and i_F respectively. We can now write, corresponding to equation (26),

$$i_F = A\beta i_1 \quad \dots\dots(29)$$

where A is the current gain of the amplifier and β the feedback ratio (alternatively A can be the transfer impedance of the amplifier and β the feedback admittance—the product $A\beta$ still being dimensionless). The result corresponding to (28) is

$$i_1 = (i_s + i_{NS} + i_{N1})R_S / (R_S + Z_i)(1 - A'\beta) \quad \dots\dots(30)$$

where $A' = AR_S / (R_S + Z_i)$

Now equations (28) and (30) will give us the closed-loop signal/noise ratio, that is the signal/noise ratio as modified by the action of the feedback. It is most convenient to compare this with the open-loop signal/noise ratio, that is the signal/noise ratio that would be obtained if the feedback parameter β were assumed to be zero. We see from equations (28) and (30) that whether the feedback is of the series or the parallel type, its action is to reduce the effect of the signal, the source noise, and the amplifier noise all in the ratio $(1 - A'\beta)$. It is important to remember that this is a complex function of frequency.

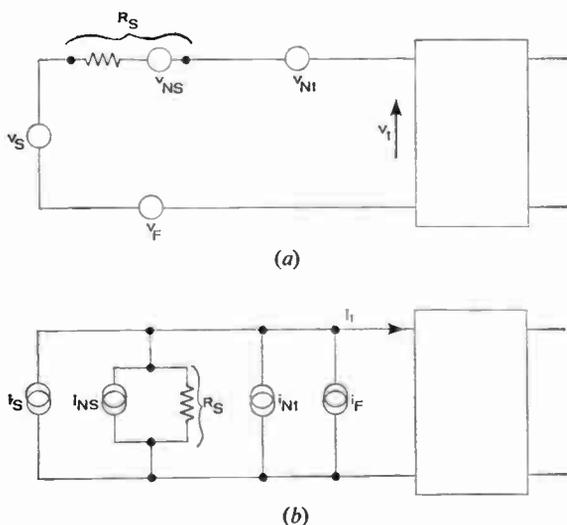


Fig. 7. Equivalent circuit of (a) series-feedback amplifier, (b) parallel-feedback amplifier. The 'black box' represents a noiseless amplifier.

It follows that if we are considering only the noise in a narrow bandwidth around the signal frequency then the closed-loop noise figure is the same as the open-loop noise figure. However, if we are considering the noise over a relatively broad band, the ratio of the closed-loop and the open-loop noise figures depends on the frequency dependence of the factor $(1 - A'\beta)$. The form of this frequency dependence also determines the relationship between the closed-loop bandwidth and the open-loop bandwidth, and the effect on the noise figure may be conveniently expressible in these terms; for instance, if (as is frequently the case) the closed-loop bandwidth is

greater than the open-loop bandwidth, then the overall closed-loop noise figure is greater than the overall open-loop noise figure.

We may simplify the discussion by supposing that under closed-loop conditions the amplifier is followed by a filter which 'tailors' the frequency response to be the same as the open-loop response. In this case the closed-loop noise figure is the same as the open-loop noise figure. We may summarize this result by the following statement: *after any changes in the frequency response have been allowed for, the closed-loop noise figure is equal to the open-loop noise figure.*

From a practical point of view, the situations described by Figs. 8 are idealized. In fact, the feedback network will have a finite impedance which appears in series (Fig. 7(a)) or in parallel (Fig. 7(b)) with the input circuit; also it will generate some noise which adds to the equivalent amplifier noise generator. It is the job of the circuit designer to make these effects negligible.

4.2. Practical Application of the Theory

It follows from the result derived in the previous section that when designing a low-noise feedback amplifier, one must design the circuit so that it has the required noise figure before the feedback loop has been closed. Suppose, for example, that we wish to build an amplifier to work from a source resistance of 100 kΩ and to have an input resistance which is very high compared with this value. By the use of series feedback, the required input resistance can easily be achieved whatever the operating current of the input transistor. In order to decide on the correct value of this current, we must consider what the situation would be in the absence of the feedback. Now as we have seen in Section 3.2, a good noise figure is not obtainable from a bipolar transistor in the common-emitter configuration unless it is operated with its input resistance substantially greater than the source resistance; it follows that in the example quoted, the common-emitter input resistance of the input transistor must be more than 100 kΩ, and the operating current must be chosen accordingly.

This example helps to make clear a fallacy which has often led to unsatisfactory circuit and system design. This is the idea that the application of series feedback to an amplifier, because it increases the input resistance, can enable the amplifier to give a satisfactory noise figure from a higher source resistance than before.

Another error which has often led to difficulties is the failure to realize that when a series feedback resistor is inserted in the emitter lead of the input transistor, this resistor is effectively in series with the signal source and its value must be added to the series noise resistance of the amplifier.

Similarly, any resistor used for parallel feedback has the effect of reducing the parallel noise resistance of the amplifier. These considerations do not, of course, affect the general result that, for a given frequency response, the closed-loop noise figure is equal to the open-loop noise figure; but in calculating the open-loop conditions we must take care to include the effects of all the components of the feedback network, even although the latter is assumed to be inoperative.

4.3. CE, CC and CB Connections

It is often said that there are three basic ways of using a bipolar transistor in a linear circuit, these being the common-emitter (CE), common-base (CB), and common-collector (CC) configurations. Now although this approach is often useful it can also be very misleading, and has certainly led to a plethora of time-wasting algebraic work. Generally speaking, we should regard the basic amplifying action of the transistor as being with the input voltage applied between base and emitter, and the output current generated in the emitter-collector circuit; this approach is clearly brought out by the well-known hybrid- π equivalent circuit. The CB and CC configurations then appear as feedback modifications of the basic action: in normal feedback terminology, the CB arrangement is one of parallel current feedback, and the CC (emitter-follower) one of series voltage feedback.⁸

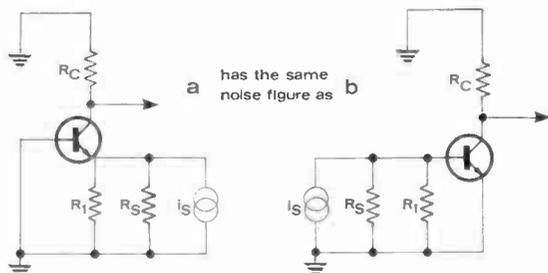


Fig. 8. CB noise figure.

The application of the theory of this section to the CB case is perfectly simple. Figure 8(a) shows the a.c. equivalent circuit of a CB-connected transistor, with an emitter resistor R_1 and operating from a signal source of resistance R_S , shown here in its current-generator equivalent form. The corresponding open-loop arrangement is shown in Fig. 8(b) and, by the principle developed in Section 4.1, we see that Figs. 8(a) and 8(b) will both give the same noise figure at any given frequency, assuming of course that the transistor is operating under the same d.c. conditions in the two cases. Some designers have made

the mistake of assuming that, because the CB arrangement has a much lower input resistance than the CE arrangement, it can be used in conjunction with a much lower source resistance without detriment to the noise figure; it should be clear from Fig. 8 that this is not correct.

The case of the CC (emitter-follower) amplifier is not so simple, because of the fact that the output is the voltage across the emitter resistor itself, rather than the output current of the transistor. When a detailed calculation is done, we find the result illustrated in Fig. 9.

Since the feedback is of the series type, the signal source is most appropriately represented as a voltage

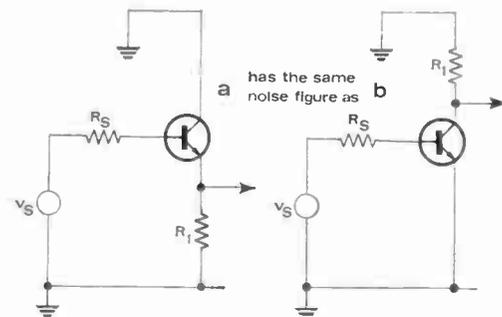


Fig. 9. CC noise figure.

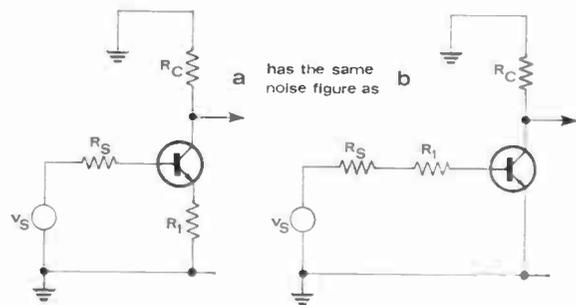


Fig. 10. Noise figure of CE amplifier with emitter feedback.

generator. The figure shows that in calculating the noise figure of a CC stage we must regard the resistor R_1 as being merely shunted across the output and not as being an additional resistance in series with the input circuit.

This is in contrast with the situation shown in Fig. 10, which is a CE amplifier with emitter feedback. Here the feedback is of the normal series current type, and the emitter resistor appears in series with the input circuit.

It is easy to make an experimental demonstration of the essential difference between the CC stage and

the CE stage with emitter feedback. The circuit of Fig. 10(a) may be set up with $R_C = R_1$, and driven from an a.f. signal source of very low resistance. With typical circuit values, say $I_C = 1$ mA and resistors of several kilohms, there is a spectacular difference between the signal/noise ratio measured at the emitter and at the collector, although of course the magnitude of the signal gain is unity at each point.

4.4. Use of Negative Feedback to Improve Noise Figure

In one sense, as we have seen, it is impossible to improve the noise figure of an amplifier by applying negative feedback; this statement is true when we are comparing the closed-loop situation with the open-loop situation for an amplifier whose circuit remains otherwise unchanged. On the other hand, we may compare one amplifier which provides certain performance parameters (e.g. gain, input impedance, output impedance, power output, efficiency, bandwidth) without the use of negative feedback, with another amplifier which utilizes the principles of negative feedback to obtain the same performance parameters. On this basis of comparison, it is often possible to obtain a great improvement in noise figure by the use of negative feedback.

A simple example of this arises in the case where one requires an amplifier to have a very low input resistance, say 1Ω , but to give a good noise figure from a comparatively high source resistance, say 1 k Ω . One could obtain the required input resistance by connecting a 1Ω resistor across the input terminals of a non-feedback transistor amplifier; but the effect on the noise figure would be disastrous. By the use of the parallel-feedback (operational-amplifier) technique we can obtain the required input resistance without any adverse effect on the noise figure.

Another example arises when we are dealing with power output stages, which may have a very poor noise figure. By adding a low-noise preamplifier, as discussed in Section 2.6, we may improve the noise figure; on the other hand, this preamplifier may increase the overall gain to an unacceptably high value. The use of negative feedback will enable the gain to be brought back to the required value without change in the overall noise figure.

Yet another case of practical importance is that in which we are required to provide a low-noise amplifier of widely variable gain. An output attenuator is usually not acceptable because of its effect on the dynamic range; and an input attenuator will ruin the noise figure. It is often possible to obtain a variable gain with constant noise figure by providing a negative-feedback network with variable feedback ratio.

Basically, the virtue of negative feedback is that it enables the input resistance, and the gain, of the

system to be adjusted to the required value without detriment to the noise figure.

5. Practical Circuit Design

5.1. Worst-case Design of the Input Stage

The discussion in the previous sections has in principle given enough information to enable low-noise a.f. amplifiers to be designed. However, in practice the designer is not usually in possession of all the required parameters, and in this section we shall consider some experimental short-cuts which will prove useful.

The basic problem is the lack of reliable information about flicker-noise parameters, although in a later paper we shall be presenting the results of a large number of flicker-noise measurements on devices submitted to us by manufacturers as low-noise devices.

From a practical engineering point of view, we want to design the input circuit in such a way that we are sure that the component values are correct. If the circuit then does not give the required noise figure, we shall know that the fault is in the input device itself rather than in its operating conditions; the selection of a suitable input device, if only by a process of trial-and-error, is then a straightforward matter.

We shall assume that the required upper-limit noise figure F_1 has been specified—this will often be in the neighbourhood of 1–3 dB, because such a noise figure is likely to be indistinguishable from unity (0 dB). The first case to be considered is that in which the source resistance R_s is specified, and we do not wish to use a transformer. As a first choice for an input device we consider a bipolar transistor, the assumption being that it is a modern device designed for low-noise work at audio frequencies, having a direct-current gain of over 100 at collector currents down to $1 \mu\text{A}$ or less. As far as flicker noise is concerned, the situation is that we cannot accurately predict the value of β'_0 (defined by equation (20)) as a function of frequency and collector current for any individual transistor; however, we know that generally the flicker-noise characteristic frequency ω_F falls as I_C is reduced, so that provided we have chosen a device in which the d.c. gain β_0 is practically independent of I_C we shall obtain the least flicker noise by operating the transistor at as low a value of I_C as possible. In practice, the choice of I_C is normally made on the basis of equation (22); although this is only correct if conditions (21) are satisfied, it is always sufficiently accurate to be used for worst-case design. Equation (22) shows that even if β'_0 were infinite a lower limit to I_C would be imposed by the requirement that

$$(r_b + r_e/2) < R_s(F_1 - 1) \quad \dots\dots(31)$$

The actual optimum value of I_C cannot be determined accurately; however, it is reasonable to assume that β'_o does not vary very rapidly with I_C so that, for a given value of β'_o , the transistor will be operating from its optimum source resistance when the second and third terms in equation (22) are equal. Now the worst acceptable value for β'_o is that which gives the required noise figure F_1 when the optimum source resistance is used; accordingly we obtain the condition

$$(r_b + r_e/2) = R_S(F_1 - 1)/2 \quad \dots\dots(32)$$

as the basis for calculating the operating current of the input transistor.

If we are dealing with broad-band amplifiers, the effective value of β'_o is a suitable average over the required frequency range; the same principles apply to the choice of operating current.

In applying equation (32), we put in typical values of r_b for the transistor type to be used, if this has already been decided on; otherwise we put in a value of 100 Ω , which is about the lowest value that can be reliably obtained from present-day low-noise devices. If equation (32) then gives $r_b \ll r_e/2$, we may use a device with a higher value of r_b without substantially affecting the noise figure. On the other hand, it may be impossible to satisfy equation (32) with $r_b = 100 \Omega$ and a positive value of r_e , in which case we use the parallel-transistor technique described in Section 2.5. In fact, it is generally undesirable to operate a transistor under conditions where $r_e < r_b$, because such conditions make unduly high demands on the value of β'_o if a good noise figure is to be achieved; in practical terms, this means that input transistors should not normally be run at collector currents greater than about 250 μA . Thus if equation (32) leads us to the requirement that $r_e < r_b$, we should consider the use of n transistors in parallel, applying equation (32) with R_S replaced by nR_S .

We now come to the case where an input transformer is to be used. In principle this means that we have a free choice of our value of effective source resistance, although in practice there will be an upper limit to the usable secondary impedance imposed by winding capacitance and also by mechanical considerations. In practice, for a noise figure of 1 dB it is not necessary to use a secondary impedance greater than 10 k Ω .

So far we have not considered the use of j.f.e.t. as an input device. This is because most of the currently available j.f.e.t.s have high values of R_{Nv} at low audio frequencies so that satisfactory noise-matching is only obtained from rather large source resistances. However, the time may come when low-cost j.f.e.t.s are available with values of R_{Nv} of a few hundred ohms throughout the audio-frequency range,

and they will then be the natural first choice as input devices.

5.2. The Feedback Resistor

Any modern circuit design is likely to incorporate a high degree of negative feedback. This will involve the inclusion of a resistor either in series or in parallel with the input circuit, and we must now discuss the way in which the results of Section 5.1 may be modified to allow for the effect of this feedback resistor.

The effect on the input circuit of a series feedback resistor R_{FS} can very simply be included in the discussion of Section 5.1 by replacing the base resistance r_b by the total effective ohmic series resistance ($r_b + R_{FS}$). In the case where we have a parallel feedback resistor R_{FP} , we must include in equation (22) an additional term (R_S/R_{FP}); generally speaking, we should choose R_{FP} so that the contribution of this term to the noise figure is less than the contributions of the terms dependent on transistor characteristics. This implies that

$$R_{FP} > 3R_S/(F_1 - 1)$$

but this criterion may, in practical cases, not be sufficiently rigorous. When the required value of R_{FP} has been established, the operating current of the first stage may be determined from equation (32) with the quantity $(F_1 - 1)$ replaced by $(F_1 - 1 - R_S/R_{FP})$. This modification places more stringent requirements on the value of β'_o , for which reason it may be desirable to satisfy the above condition with as large a margin as is practicable.

5.3. Worst-case Design of the Second Stage

One of the commonest errors in low-noise circuit design is the failure to ensure that the second stage does not contribute to the total flicker noise. In this section we shall make a worst-case analysis which will lead to a rule-of-thumb to aid the circuit designer. The worst case can be set up by assuming that the source resistance is zero, so that the input transistor is generating less noise than it ever will in a practical situation. With this assumption, we shall attempt to find a criterion which ensures that the noise contribution of the second stage is negligible.

If the input transistor is being operated under low-noise conditions, the effect of the base resistance r_b can be neglected in this discussion. From equations (23) we see that the noise in the input transistor can therefore be represented by a voltage generator with mean-square value $2r_{e1}kT\Delta f$, where r_{e1} is the emitter resistance of the input transistor. Now the transfer admittance (mutual conductance) of this transistor is approximately $1/r_{e1}$, so the noise current generated by the input transistor in the CE or CB configuration

has a mean-square value $2kT\Delta f/r_{e1}$. Now, as we shall see shortly, the noise contribution of the second transistor is normally dominated by the effect of the equivalent noise-current generator. Equations (23) show that the mean-square value of this generator is $2kT\Delta f/\beta'_{o2}r_{e2}$, where the subscript 2 refers to the values for the second transistor. If this is not to contribute substantially to the overall noise, we must establish the condition that

$$2kT\Delta f/\beta'_{o2}r_{e2} \ll 2kT\Delta f/r_{e1}$$

which is equivalent to the condition

$$\beta'_{o2}r_{e2} \gg r_{e1} \quad \dots\dots(33)$$

To obtain the corresponding condition for the case of a CC (emitter-follower) input stage, we observe that if R_s is zero the output of the first transistor is equivalent to a voltage generator of mean-square value $2r_{e1}kT\Delta f$ in series with a resistance r_{e1} . This is equivalent to a current generator of mean-square value $2kT\Delta f/r_{e1}$ in parallel with r_{e1} , and it follows that condition (33) is also applicable to the CC configuration.

Although equation (33) has been derived on the assumption of a somewhat extreme worst-case, it provides a very convenient design basis for the instrumentation or communications engineer. At worst, it will result in a circuit which contains one more transistor than is strictly necessary; but transistors are cheap, and design and development effort is expensive.

It is interesting to notice that equation (33) cannot possibly be satisfied either by the traditional 'Darlington pair' configuration, in which the emitter current of the input transistor is equal to the base current of the second transistor, or by the analogous complementary design in which the collector current of the input transistor is equal to the base current of the second transistor (see Fig. 11). This is because the collector current in the second transistor is greater than that in the first transistor by a factor of β_{o2} , so that $r_{e2} = r_{e1}/\beta_{o2}$. Thus with a zero source resistance, even in the absence of flicker noise, the second transistor makes a noise contribution equal to that in the first transistor; in practice the situation will probably be made very much worse by the effect of flicker noise. The direct-coupled stages shown in Fig. 11 can be greatly improved by shunting the input of the second transistor with a resistor to enable it to run with a lower value of collector current, appropriate changes being made in the remainder of the circuit.

Now even a reasonably low-noise transistor may show a value of β'_o approaching unity at the lower end of the audio-frequency range; this would be

true, for example, at 20 Hz if β_o were 100 and the flicker-noise characteristic frequency ω_F were 2 kHz. Therefore, if we want to be reasonably sure, even under the worst conditions (zero source resistance) and in the lowest part of the audio-frequency range, that the second transistor does not contribute to the overall noise, we make use of condition (33) in the form

$$r_{e2} = r_{e1} \quad \dots\dots(34)$$

This condition is equivalent to the statement that the d.c. collector current in the second transistor should be equal to that in the first transistor.

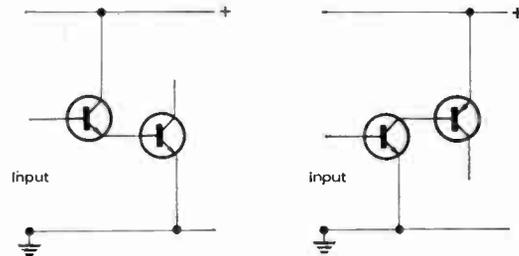


Fig. 11. Examples of noisy circuit design.

As a last step, we must arrange that the voltage gain of the first stage is sufficiently high to ensure that we have satisfied the requirement, mentioned above, that the voltage-generator noise in the second transistor should be negligible. First we assume that the first stage has the CE or CB configuration, with collector load resistance R_C —this being assumed to include the effect of the input resistance of the second transistor. As we have seen, the noise current generated by the input transistor has a mean-square value $2kT\Delta f/r_{e1}$, and this provides a noise voltage across R_C with mean-square value $2kT\Delta fR_C^2/r_{e1}$. Now the voltage-generator noise in the second transistor has a mean-square value $2r_{e2}kT\Delta f$, so our required condition is

$$R_C^2 \gg r_{e1}r_{e2} \quad \dots\dots(35)$$

This condition is easily satisfied. The lowest value we are likely to use for R_C arises in a direct-coupled complementary circuit where the direct voltage drop across R_C is only about 600 mV. Assuming for example that the d.c. collector current in the first transistor is 10 μ A, then r_{e1} will be 2.5 k Ω and R_C will be 60 k Ω . It is clear that (35) will be satisfied even if the second transistor has the same d.c. collector current as the first so that $r_{e2} = r_{e1}$.

For a CC (emitter-follower) input stage the situation is not so favourable. Since the voltage gain is unity, the condition equivalent to (35) is

$$r_{e1} \gg r_{e2} \quad \dots\dots(36)$$

which can be formally obtained from (35) by putting $R_C = r_{e1}$ but which is really self-evident. Obviously, we cannot satisfy both condition (34) and condition (36); for this reason the CC configuration is inherently more noisy for an input stage than the CE or the CB configurations. In some practical cases the difference is so slight as to be negligible, and the CC stage may be preferred for other reasons.

5.4. General Comments on Circuit Design

There are many textbooks on transistor circuit design, but it is an unfortunate fact that on the whole they are highly unsatisfactory.

We must avoid being misled by the idea of a 'norm' of amplifier design consisting of a string of separate single-transistor or two-transistor stages with capacitor coupling and with feedback (if any) provided by individual emitter resistors. The amplifier should be designed as a whole, with direct coupling between the transistors unless there is some good reason to the contrary, and with a high degree of d.c. and a.c. feedback taken around each direct-coupled group. As a matter of course, silicon transistors will be used throughout. Apart from the obvious economies in components, this approach enables a greater number of transistors to be included in the feedback loop so that the noise contribution of the feedback resistor can be reduced to the lowest possible level.

We must also avoid the assumption that noise considerations are a sort of 'extra' which only needs to be taken into account under exceptional circumstances. As a matter of general engineering workmanship, the good designer ensures that, whatever the specification, his circuits do not show an *unnecessarily* high noise level; and a prospective purchaser who has

to choose between two otherwise identical pieces of equipment will always choose the one showing the lower noise level. It is by no means unusual to find a factor of 100 or more between the noise levels shown by the product of a 'noise-conscious' designer and the product of a designer who follows typical textbook principles.

6. Acknowledgment

This work was supported by a Ministry of Defence contract.

7. References

1. Faulkner, E. A., 'Optimum design of low-noise amplifiers', *Electronics Letters*, **2**, pp. 426-7, November 1966.
2. van der Ziel, A., 'Theory of shot noise in junction diodes and junction transistors', *Proc. I.R.E.*, **43**, pp. 1639-46, November 1955.
3. van der Ziel, A., 'Theory of shot noise in junction diodes and junction transistors', *Proc. I.R.E.*, **45**, p. 1011, July 1957. (Letter.)
4. Faulkner, E. A. and Dawnay, J. C. G., 'Characteristics of silicon transistors', *Electronics Letters*, **1**, pp. 224-5, October 1965.
5. van der Ziel, A., 'Thermal noise in field-effect transistors', *Proc. I.R.E.*, **50**, pp. 1808-12, August 1962.
6. Knott, K. F., 'Conditions for optimum noise performance in l.f. amplifiers employing junction f.e.t.s.', *Electronics Letters*, **4**, p. 92, March 1968.
7. Knott, K. F., 'Comparison of varactor-diode and junction-f.e.t. low-noise l.f. amplifiers', *Electronics Letters*, **3**, p. 512, November 1967.
8. Faulkner, E. A., 'Principles of Linear Circuits'. (Chapman & Hall, London, 1966.)

Manuscript received by the Institution on 2nd April 1968. (Paper No. 1197/CC12.)

© The Institution of Electronic and Radio Engineers, 1968

Some Measurements on Low-noise Transistors for Audio-frequency Applications

By

E. A. FAULKNER†
AND
D. W. HARDING‡

Summary: The device parameters which are relevant in the design of low-noise circuits are discussed and the results of experimental work on a number of commercial types are presented.

1. Introduction

This contribution is intended to be read as a supplement to a paper‡ on the design of low-noise audio-frequency amplifiers, and the equations referred to here are those given in that paper.

2. Relevant Parameters

The noise parameters of a bipolar transistor are highly dependent on the d.c. collector current, and the key to successful bipolar circuit design is a correct choice of the operating currents of the transistors in the initial stages. In practice, these currents may lie anywhere in the range 100 nA–1 mA, and it is therefore desirable to have devices available which have satisfactory noise characteristics over this whole range. From equations (20), (21), and (22) we see that the relevant parameters are:

(a) The direct current gain β_o . The noise figure actually depends on the quantity $\beta'_o = \beta_o/(1 + \omega_F/\omega)$, and we see from equations (23) and (11) that a value of $\beta'_o = 50$ is necessary at a given frequency if a minimum noise figure of 0.6 dB is to be obtainable. Such a minimum noise figure would provide a satisfactory margin for meeting a 1 dB specification, allowing for non-optimum operating conditions, provision of feedback and protection resistors, noise contributions from later stages, etc. Since β'_o cannot be greater than β_o , we may put $\beta_o = 50$ as a reasonable minimum target figure for a general-purpose low-noise transistor, down to collector currents lower than 1 μ A. On the other hand, we must remember that a transistor may still have a substantial low-noise capability ($F_{\min} < 1.5$ dB) down to $\beta_o = 10$.

(b) The flicker-noise characteristic frequency ω_F . The significance of this parameter is different according to whether we are dealing with spot-frequency or broad-band noise figures.

Working at a spot frequency ω the important figure is the value of β'_o at the frequency ω ; from equation (20) we may write

$$1/\beta'_o = 1/\beta_o + \omega_L/\omega$$

where

$$\omega_L = \omega_F/\beta_o$$

The parameter ω_L , which may be called the *lower limiting frequency*, is a more useful one than ω_F for characterizing the 'flicker-noisiness' of a transistor; it is the frequency at which β'_o has fallen to unity (we assume that $\beta_o \gg 1$) and at which the transistor has therefore lost any claim to low-noise capability. Assuming that $\beta_o > 50$, we can calculate that at frequencies greater than $10\omega_L$ it is possible to obtain a noise figure better than 1.5 dB. We shall see shortly that a good low-noise bipolar transistor shows $\omega_L < 10$ Hz.§

In a broad-band application, we require to know the effective average value of β'_o , and in order to obtain this it is necessary to integrate the mean-square noise current in equation (18) over the required frequency band, say ω_1 to ω_2 . The result, expressed in terms of $(\beta'_o)_{av}$, the effective average value of β'_o , is:

$$1/(\beta'_o)_{av} = 1/\beta_o + [\omega_L/(\omega_2 - \omega_1)] \log(\omega_2/\omega_1)$$

Now in a typical audio-frequency application we might have ω_1 and ω_2 equal to 10 Hz and 10 kHz respectively; putting ω_L equal to 10 Hz we find that the last term in the above equation is less than 1/140. In this case we can obtain $(\beta'_o)_{av} = 50$ for $\beta_o = 80$, and it is clear that flicker noise will not have a dominant effect on the broad-band noise figure. Even if ω_L is as high as 100 Hz, the last term is still only 1/14 and we obtain $(\beta'_o)_{av} > 10$ for $\beta_o = 50$.

(c) The base resistance r_b . The value of this parameter to be used in noise-figure equations is not necessarily equal to the high-frequency value sometimes quoted by manufacturers, and must be obtained from noise measurements. In practice we find that

§ When we specify the angular frequency in Hz, we are taking 1 Hz to be equal to 2π rad/s.

† J. J. Thomson Laboratory, University of Reading, Whiteknights, Reading, Berkshire.

‡ Faulkner, E. A., 'The design of low-noise audio-frequency amplifiers', *The Radio and Electronic Engineer*, 36, No. 1, pp. 17–30, July 1968.

Table 1
Experimental results of measurements on low-noise bi-polar transistors

Transistor type	Polarity	β_o ($I_C = 100 \mu A$)		β_o ($I_C = 1 \mu A$)		ω_L (Hz) ($I_C = 100 \mu A$)		ω_L (Hz) ($I_C = 1 \mu A$)		r_b (Ω)		ω_T (kHz) ($I_C = 1 \mu A$)
		worst	average	worst	average	worst	average	worst	average	worst	average	
Texas 2N3707	n-p-n	200	250	88	150	44	17	6.9	2.1	300	200	750
Texas 2N4058	p-n-p	110	200	28	105	10	2.1	2.3	1.0	210	170	250
Texas 2S502	n-p-n	130	230	78	120	>260	19	13	1.4	620	180	600
Mullard BC109	n-p-n	200	300	110	200	10	4.4	7.6	1.8	430	360	500
Mullard 2N930	n-p-n	150	230	63	130	260	56	260	9.3	490	380	300
S.G.S.-F. BFX37	p-n-p	170	190	140	170	21	3.3	1.2	0.6	75	60	350
S.G.S.-F. BFY77	n-p-n	150	190	82	94	140	51	12	5.5	620	400	600
Motorola 2N4126	p-n-p	110	190	41	83	18	4.3	7.3	1.9	60	40	300
Motorola MPS6555	n-p-n	140	150	43	54	35	17	20	7.5	990	860	80
Sprague 2N4384	n-p-n	160	240	95	140	14	4.8	4.5	0.8	210	130	200
Sprague 2N4413	p-n-p	160	240	95	170	35	15	50	3.4	100	85	200
Ferranti 2N930	n-p-n	130	150	51	76	7.9	3.1	2.3	1.2	390	230	400

devices with the lowest values of r_b do not necessarily have the best noise performance in other respects, and it is desirable to have devices available which have been chosen for low r_b as well as devices which have been chosen for high β_o and low ω_L .

(d) The value of the transition frequency ω_T at the lowest value of collector current likely to be used. Our whole discussion has been on the basis, stated in Section 3.1 of the main paper, that we are considering frequencies well below $\omega_T/\sqrt{\beta_o}$. If the frequency range extends to 20 kHz, this condition implies that ω_T for the transistor must be greater than $20\sqrt{\beta_o}$ kHz, or say 200 kHz for β_o equal to 100. Now this may not seem very high when we are dealing with silicon planar transistors for which the manufacturers quote values of 30 MHz upwards at collector currents of 1 mA; but ω_T falls off as collector current is reduced, and in evaluating a transistor as a general-purpose low-noise device it is worth while to check that it satisfies the criterion that ω_T exceeds say 200 kHz at collector currents down to 1 μA .

3. Manufacturers' Specifications

If we look down a list of manufacturers' specifications of small-signal low-noise devices, we are unlikely to find any parameter specified at collector currents below 100 μA , a very common value being 1 mA. This has led many circuit designers, and many authors of texts on circuit design, to assume that such devices are 'intended' for use at these currents, and to sacrifice considerations of transistor physics to these 'intentions' of the manufacturers.

An additional complication is introduced by the fact that manufacturers tend to designate some of their transistors as 'low-noise' devices, and then to quote audio-frequency noise figures in the region 3-6 dB, very much worse than the actual minimum noise figure of a low-noise device.

For these reasons we have undertaken a programme which involved asking a number of transistor manufacturers to send us samples of devices which they regard as suitable for low-noise applications in the audio-frequency region. We have carried out noise measurements at room temperature on these devices, the results of which are described in the next section.

4. Experimental Results

The experimental results are summarized in Table 1. The current-noise parameters were measured at collector currents of 1 μA and 100 μA , the base resistance at 1 mA, and the transition frequency at 1 μA . Six transistors of each type were measured, and the 'average' figures given in Table 1 refer to the arithmetic mean of the five best values.

The most remarkable feature of these results is that, in every single case, the noise parameters of the 'low-noise' transistors were very much better than those which might reasonably be deduced from the manufacturers' published data. We have every reason to be confident that the devices were not specially selected.

As a specific example we may quote the Texas 2N3707, for which the manufacturers claim a value

of $\beta_o > 100$ at $I_C = 100 \mu A$, and a noise figure of < 5 dB. Our measurements showed average values of 250 and 150 for β_o at $100 \mu A$ and $1 \mu A$ respectively, and the minimum noise figure of all the specimens tested was better than 0.5 dB. Similar figures were given by a random selection of devices obtained from normal commercial sources.

5. Conclusions

The results quoted here are taken over a relatively small number of specimens of each type, and there is no implication that the averages quoted are accurately representative of the production batch from which the devices came. On the other hand, it is doubtful whether an extended statistical analysis over a large number of specimens would be significantly more useful in practice to the circuit design

engineer, because of variations between individual batches.

With these limitations borne in mind, the results given in Table 1 have proved to be extremely useful in the selection of suitable transistor types for low-noise equipment. Naturally, the final choice made by the user will depend on considerations of cost, availability and environmental range as well as the parameters given in the Table.

6. Acknowledgments

The authors are greatly indebted to Viscount Downe for his contribution to this work. Financial support was provided by a Ministry of Defence contract.

Manuscript received by the Institution on 2nd April 1968. (Contribution No. 105/CCI3.)

© The Institution of Electronic and Radio Engineers, 1968

Letters to the Editor

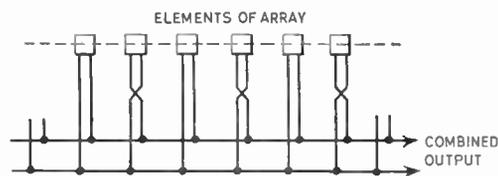
Superdirective Arrays

SIR,

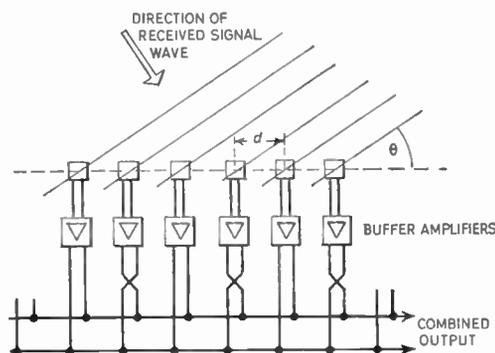
I read the paper on superdirective arrays by Professor Tucker in the October 1967 issue of *The Radio and Electronic Engineer* with considerable interest, since I have felt for some time that it should be possible to exploit superdirectivity to a greater extent than is done at present, and have had a number of discussions with my colleagues on this subject. In some of these discussions we have tended to assume that buffer amplifiers might be used, and as a result I would like to make some comments on points arising from Professor Tucker's paper.

Element Interaction. If we restrict attention to an array having equally spaced elements with weighting factors of ± 1 , and with equal electrical paths between the elements and the output (i.e. as in Fig. 2(a) of the paper, but assuming the bus-bars to be of zero electrical length), then I believe the element interaction problem may be less serious than appears at first sight. I suggest that, as a first approximation, the elements voltages for the array of the sort given in Fig. 2(a) can be considered as made up of following two contributions:

- (i) The voltage distribution resulting from the incident field with the output terminals short-circuited.
- (ii) The voltage distribution that results when in the absence of an external field, the output terminals are driven by a voltage equal to the output voltage existing under operating conditions.



(a) With direct electrical coupling.



(b) With buffer amplifiers.

Fig. 2. Superdirective array connections.

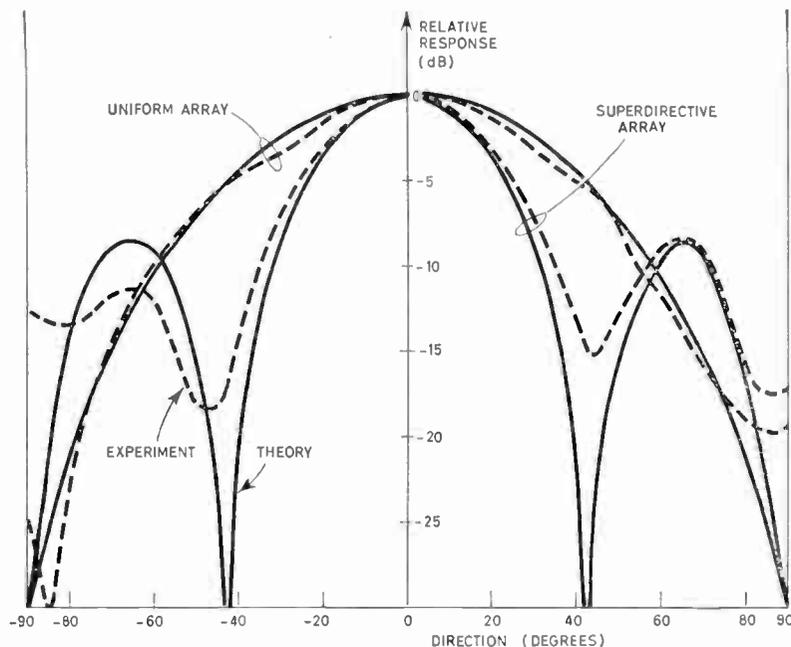


Fig. 3. Comparison of experimental and theoretical results for a 420 MHz array of three radio dipoles spaced at quarter-wave intervals. Buffer amplifiers were used as in Fig. 2(b). The superdirective array merely had the phase of the output from the central dipole reversed before adding to the others.

For a non-resonant superdirective array, (ii) should be much smaller than (i), since we know that in these circumstances the output voltage is normally small compared with that in the individual elements, thus the voltage distribution will be determined mainly by (i). This has the same form as given by Professor Tucker for the array in Fig. 2(b), in which elements are independent except for coupling through the field and are closed on identical impedances, but the values of the m are likely to be different, of course.

The m are transfer coefficients and involve the ratios of self to mutual impedance for the elements. The use of amplifiers should enable us to avoid complications due to low values of self-impedance resulting from resonance. The mutual impedance can, of course, be reduced by adopting 'inefficient' elements (such as small hydrophones or short non-resonant dipoles). In some cases this could be made easier by use of buffer amplifiers, but the importance of noise in most superdirective systems suggests that this will not always be the case.

I believe that the real advantage of buffer amplifiers shows up when it is desired to adopt complex weighting factors for the elements. In this case we can no longer determine the voltage distribution by superimposing conditions (i) and (ii) above, since the elements do not, in general, see identical impedances, and thus the mutual impedance between elements enters into the determination of the weighting networks. This can be avoided when buffer amplifiers are employed, and in any case, as pointed out by Professor Tucker, the existence of buffer amplifiers is likely to simplify the design of the weighting network.

Bandwidth. I agree with Professor Tucker that the sharp tuning of the superdirective array has probably been overemphasized in the literature; the large range of field of the tuned array (reference 4 of the paper) seems necessary if the array is to interact with, or extract energy from, a substantial volume of space, i.e. if supergain (in the sense the term 'gain' is used in the paper on page 254) is the aim, but an inefficient array could still be superdirective.

There is, however, one possible limitation on bandwidth which I believe needs some further thought. Arrays using uniform spacing and weighting factors of ± 1 generally seem to show high side-lobes (e.g. Fig. 3 of the paper) but these can be reduced by adopting more suitable weighting factors. Although I have not seen any detailed investigations, various published results do tend to suggest that for good side-lobe performances these weights vary quite rapidly with frequency. It should, I believe, be possible to design networks—possibly of multi-path type—having the required amplitude and phase characteristics over at least reasonable bandwidths, and this task is clearly greatly eased by the insertion of buffer amplifiers which standardize the impedance conditions in which the network must operate.

Amplifier Tolerances. The output of a superdirective array is basically a small difference between large quantities, and the directional information is contained in rather small phase changes between the elements. These factors seem to imply that the buffer amplifiers will have to be matched to close tolerance in amplitude and phase. How critical this is likely to be seems to require further study.

Amplifier Design. It seems that the exploitation of superdirectivity will often depend on the availability of suitable low-noise amplifiers. There have, of course, been important advances in this field for the higher radar frequencies, but the sonic frequency range seems less well served. Superdirectivity would seem likely to be of the greatest importance at the lower frequencies, where array size tends to get out of hand, but this is just the region in which 'flicker' noise begins to become troublesome in semiconductors.

I have for some time wondered whether it would be possible to develop a parametric amplifier for this frequency region, using a mechanically-varied capacitor as pumping element. One possibility I have had in mind would be to use one of the commercially available capacitance modulators intended for use as input elements to d.c. amplifiers. I have not investigated this in any detail, and I do not know whether the commercially available components would permit the design of an amplifier of useful gain with a practical impedance level, but it would seem worth considering.

P. G. REDGMENT, M.A., C.ENG., M.I.E.E.

Admiralty Underwater Weapons Establishment,
Portland, Dorset. 24th April 1968.

The views expressed above are those of the writer and should not be taken as representing the policy of the Ministry of Defence.

SIR,

I write in connection with Professor Bell's letter¹ which comments on the paper by Professor Tucker² about superdirectivity.

Professor Bell states 'In an aerial array which consists entirely of interconnected passive elements which are fed from a single input port (pair of terminals), superdirectivity and supergain are synonymous and difficult to achieve'. I think that this is not strictly true and needs further qualification concerning the impedance matching of the array. Figure A shows a very simple form of superdirective array which is fed at a single port with a transmission line. The aerial elements are shown coupled to the line by high series impedances so that the phase of the wave travelling down the line is not disturbed sufficiently to upset the operation of the array. The currents in the elements will be very low compared with the current in the line but they will radiate and form a cardioid pattern with one null in the direction shown. The null position will be independent of frequency since the field from the two elements will arrive at this angle in the far-field with the same time delay but with opposite phase angle due to the 180° physical rotation in the transmission line. It can be shown that the directional response, with omnidirectional aerials, is given by

$$\sin \left[\frac{\pi d}{\lambda} (\sin \theta + 1) \right]$$

This is plotted against $\sin \theta$ in Fig. B for a case where the element spacing is less than $\lambda/4$. Note that the major

response lies outside of the range of real angles indicating that the array is superdirective. The array shown in Fig. A is therefore a wide-band superdirective array of interconnected passive elements with a single input port. The array is not a supergain array (most of the input power is lost in the matched load terminating the transmission line) and so supergain and superdirectivity are not synonymous. This is not a special case but is the simplest form of a class of passive superdirective arrays with the aerial elements loosely coupled to the feeder. In connection with this type of array it should be noted that it is not always necessary to insert a high impedance between the aerial elements and the feed. In practice, if the array is to be superdirective, the physical size requirement usually demands that the elements should be electrically small. For the case of small dipoles the input impedance of each dipole will be predominantly its high self reactance in series with a low radiation resistance. The current in each element will be retarded by this self reactance, but if all elements are the same length the self reactances will be equal (but frequency-dependent) and therefore the element currents can retain the correct relative amplitude and phase, over a wide frequency range, for the required directional pattern.

With the circuit as shown in Fig. A the element impedances must be very high in order that reflections from the dipoles do not upset the operation of the array.

This limitation can be overcome by using either buffer amplifiers or the arrangement shown in Fig. C. With the latter, reflections from a mismatch at one element will not affect the current in the other element as the ports A and B of the hybrid are isolated. However, since mutual

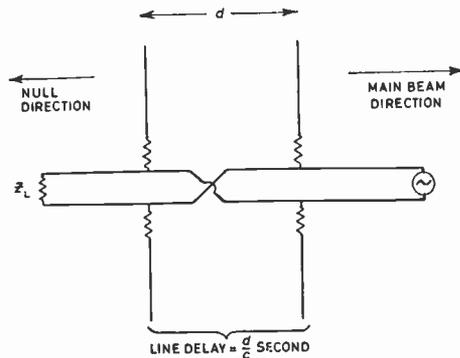


Fig. A. A simple superdirective array.

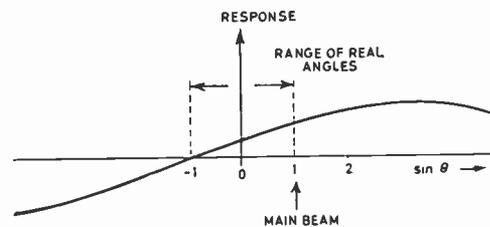


Fig. B. Directional response of array in Fig. A for $d < \lambda/4$.

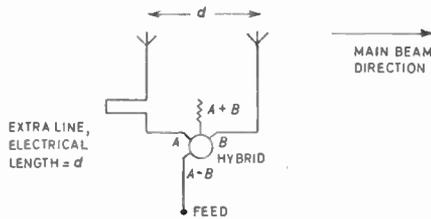


Fig. C. Alternative form of a simple superdirective array.

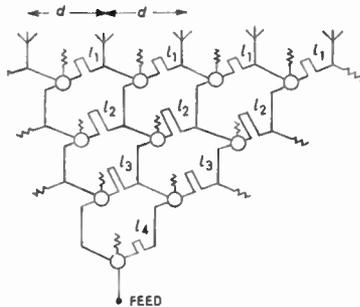


Fig. D. More general form of the arrangement shown in Fig. C.

coupling will cause the radiation resistances of the elements to differ it will still be necessary to arrange for a high impedance to interpose the element and the feed. This can be accomplished as before by using the self-reactance of shortened elements or by physically including an impedance at the element terminal.

To summarize then, it has been indicated that impedance mismatching can enable the currents required for a superdirective array to be maintained over a wide band. The problems of multiple reflections from the mismatches can be eliminated by using hybrid networks as suggested here, or by using buffer amplifiers as suggested by Professor Tucker. It is evident that supergain and superdirectivity are only synonymous if the array is lossless. Furthermore, superdirectivity appears to be readily obtainable over a wide band using passive or active circuits of directly or indirectly connected elements.

As an added point of interest Fig. D shows a more general case of the arrangement shown in Fig. C. It can be shown that zeros in the directional pattern of this array occur at angles given by

$$\theta = \sin^{-1} \left(\frac{l_m}{d} \right)$$

and are independent of frequency. By a suitable choice of the delay length, l_m , it is possible to synthesize directional patterns with constant beamwidths, and if d is much less than $\lambda/2$ the array can be superdirective.

This beam-forming matrix is an improvement of the scheme, first proposed by Dr. D. E. N. Davies³, for reducing interference from specific directions.

I. D. LONGSTAFF, PH.D.

The M-O Valve Company Limited,
Research Laboratories,
G.E.C. Hirst Research Centre,
Wembley, Middlesex.

27th May 1968.

SIR,

In view of Dr. Longstaff's remarks I should have extended my definition by adding the word *efficient*, but it has always seemed to me axiomatic that an array used for electromagnetic radiation should radiate most of the energy fed into it. This is not practicable for very-low-frequency aeriels, but neither are arrays practicable at very low frequencies. (I believe it may also be impracticable for acoustic arrays owing to the low efficiency of electro-acoustic transducers, but I was considering only electromagnetic radiators.) An array with buffer amplifiers clearly falls outside my definition and the use of resistive elements to perform a similar function is in my view equally inadmissible, though I did not explicitly exclude it.

D. A. BELL,
M.A., PH.D., C.ENG., F.I.E.E.

Department of Electronic Engineering,
The University of Hull.

4th June 1968.

Editorial Note: Professor D. G. Tucker has signified his agreement with this correspondence.

References

1. Bell, D. A., 'Superdirective arrays', *The Radio and Electronic Engineer*, 35, No. 1, p. 53, January 1968.
2. Tucker, D. G., 'Superdirective arrays: the use of decoupling between elements to ease design and increase bandwidth', *The Radio and Electronic Engineer*, 34, No. 4, pp. 251-5, October 1967.
3. Davies, D. E. N., 'Independent angular steering of each zero of the directional pattern for a linear array', *I.E.E.E., Transactions on Antennas and Propagation*, AP-15, No. 2, pp. 296-8, March 1967.

Radar Video Sweep-time Reduction

By

W. J. DUNN,

C.Eng., M.I.E.R.E.,†

E. M. SAWTELLE†

AND

A. J. MYURA†

Summary: Alpha- numerics are used to identify broadband radar targets on a p.p.i. display for air traffic controllers. A technique for writing the broadband data is described which reduces the time required to a constant time independent of radar display range. This constant time can be chosen so that adequate time is allotted for the writing of radar target associated alpha-numeric data. For alpha- numerics to be usable some method of association must be provided between the alpha- numerics and the radar target. The digital conversion techniques will allow for target detection outside the time domain.

1. Introduction

This paper describes a technique for use in air traffic control systems where alpha- numerics are written on a radar p.p.i. display. The demand for a number of characters to be written in the p.p.i. sweep display dead-time lead to the development of the technique known as 'display sweep-time reduction'.‡

If the time period available between the end of a range sweep and the beginning of the next range sweep is accepted as the time available to write alpha- numerics then the time available becomes shorter as the range gets longer. The longer the range displayed the greater the number of targets available for identification tags and the faster characters must be written. This dilemma is quantitatively illustrated for a sixty-mile range air traffic control radar whose characteristics are listed in the Appendix. The available sweep to sweep dead-time is here listed as 91 μ s. 'State of the art' deflection systems operating with these particular displays jump three-quarters of a display diameter and are fully settled to 0.1% in approximately 14 μ s. Characters can then be written at a speed of 2 μ s per character. These are requirements that tax the 'state of the art' and are expensive. By increasing the sweep-to-sweep time and by making this time independent of radar range, e.g. the quoted 91 μ s period becomes 700 μ s for which modest character writing speeds become adequate and economical to realize.

There are two approaches to this end. One is an analogue conversion and the other a digital conversion.

2. Analogue Conversion

The analogue approach to sweep-time reduction has been useful to study the attributes of the technique.

† Federal Aviation Administration, National Aviation Facilities Experimental Center, Atlantic City, New Jersey, U.S.A.

‡ U.S. Government is applying for a patent on the technique.

The timing relations of sweeps for a c.r.t.-vidicon system is shown in Fig. 1. A laboratory system as shown in Fig. 2 has been operated in accordance with the timing diagram. This provides a sweep reduction ratio of 5 : 1. Radar video from a 2.5 MHz bandwidth video amplifier is applied to the grid of the 5-inch high-resolution cathode-ray tube. The beam is deflected in a line across the diameter with single axis deflection. The deflection sweep voltage is initiated with radar trigger at time zero and continues to 60 miles or approximately 750 μ s.

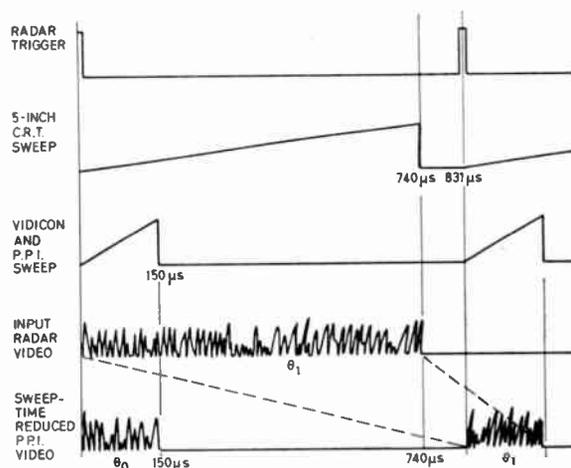
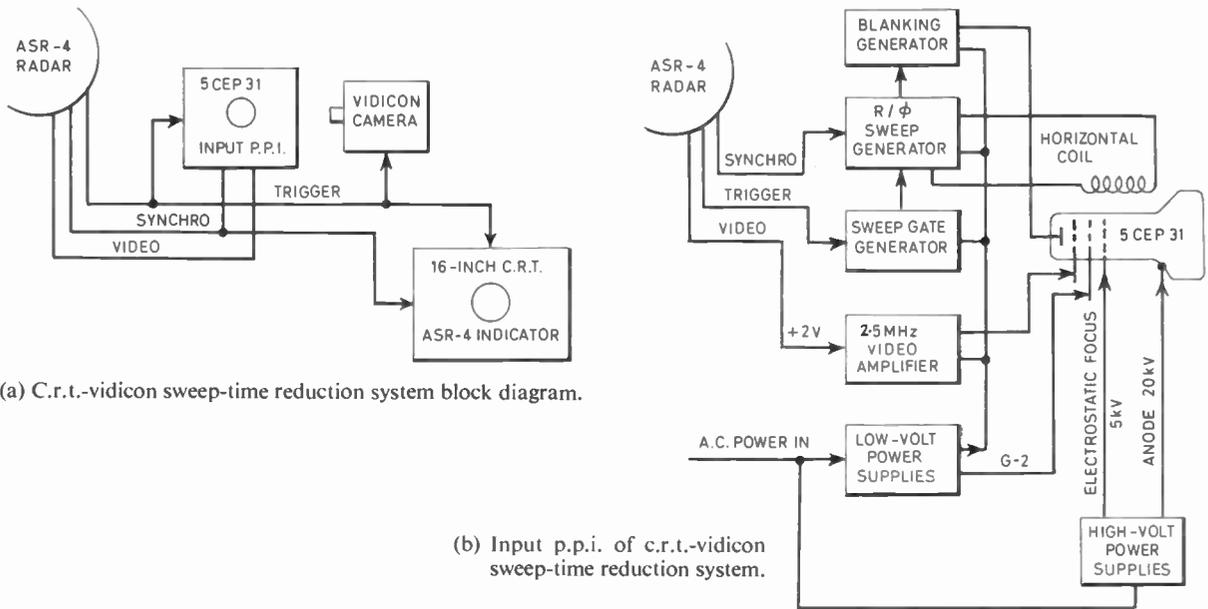


Fig. 1. Timing relations of sweeps for a c.r.t.-vidicon system.

The vidicon camera of Fig. 2 optically records the c.r.t. output. As the phosphor of the c.r.t. is activated the image of the phosphor is stored on the vidicon until read-out. The read-out 'wipes' the stored image down to approximately 50% of its initial value. A second read-out reduces the image nearly to the noise level of the vidicon. Read-out of the vidicon is accomplished in 150 μ s. As may be noted from



(a) C.r.t.-vidicon sweep-time reduction system block diagram.

(b) Input p.p.i. of c.r.t.-vidicon sweep-time reduction system.

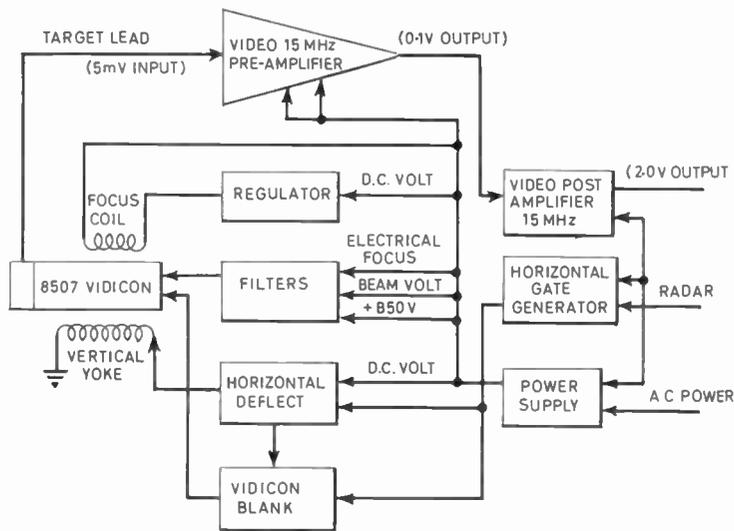


Fig. 2. (c) Vidicon camera of c.r.t.-vidicon sweep-time reduction system.

Fig. 1, the read-out is started at range zero time. The read-out start-time is largely independent of the input radar video timing. Start at time zero with radar trigger was selected because of convenience of operating without delaying the trigger by a not too stable multivibrator. The video signal having passed through the vidicon preamplifier is fed to the final output video amplifier of approximately 10 MHz bandwidth and then to the grid of the radar console

indicator c.r.t. Sweep-time reduction has been completed in a constant delay time of one radar trigger. Consequently, no error in azimuth has occurred. The video signal is applied to produce a p.p.i. display on the indicator. Photographs of the results are shown in Figs. 3 to 7. Comparable unreduced sweep-time video is shown in Figs. 8-12. The indicator c.r.t. used a P7 phosphor and the radar trails are not present since the film was sensitive to blue flash only.

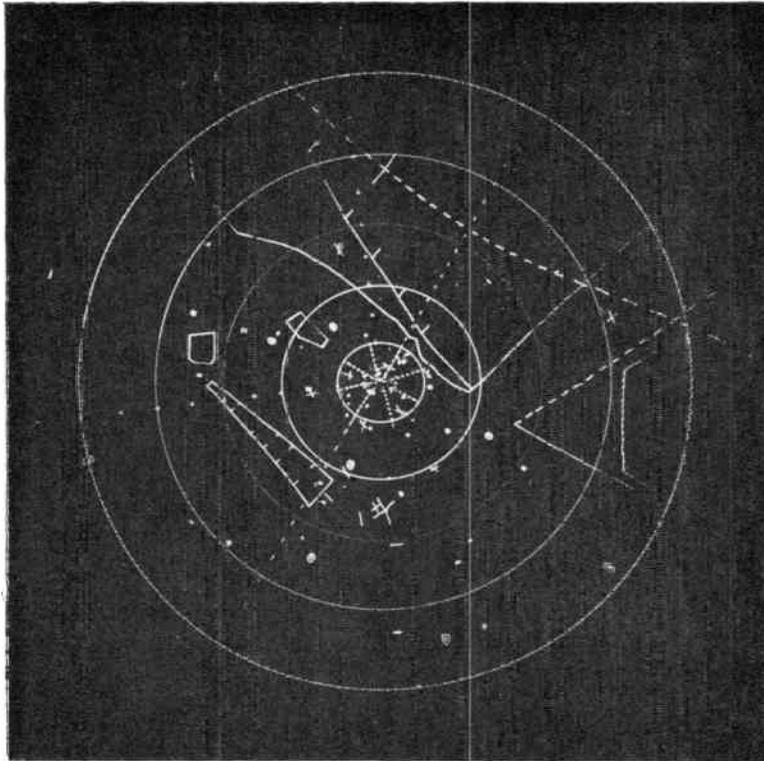


Fig. 4. 60-mile range, m.t.i. video, 10-mile r.m.

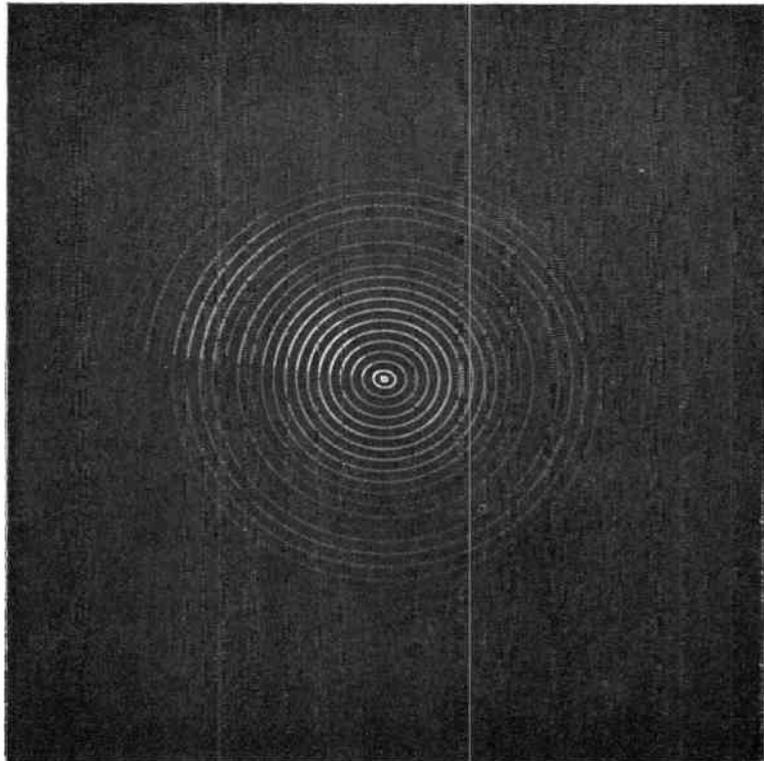


Fig. 3. 60-mile range, dual range marks (r.m.).

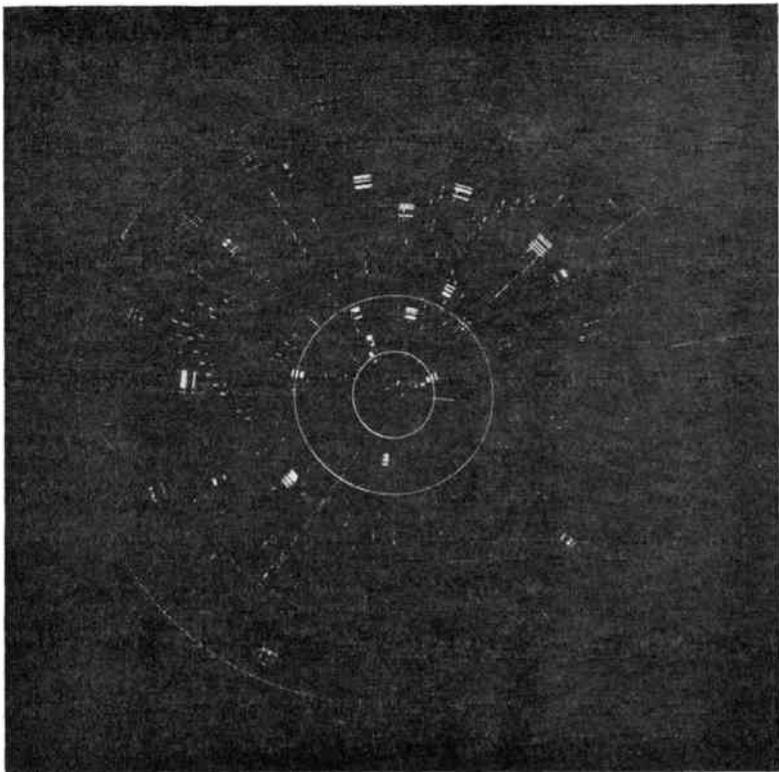


Fig. 6. 60-mile range, beacon video, 10-mile r.m.

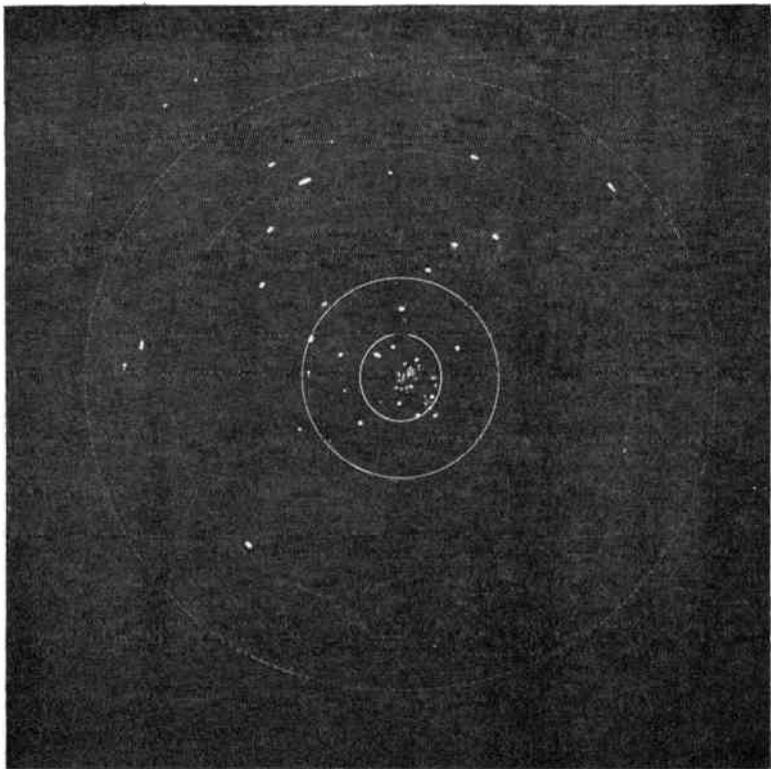


Fig. 5. 60-mile range, m.t.i., video, 10-mile r.m., sweep.

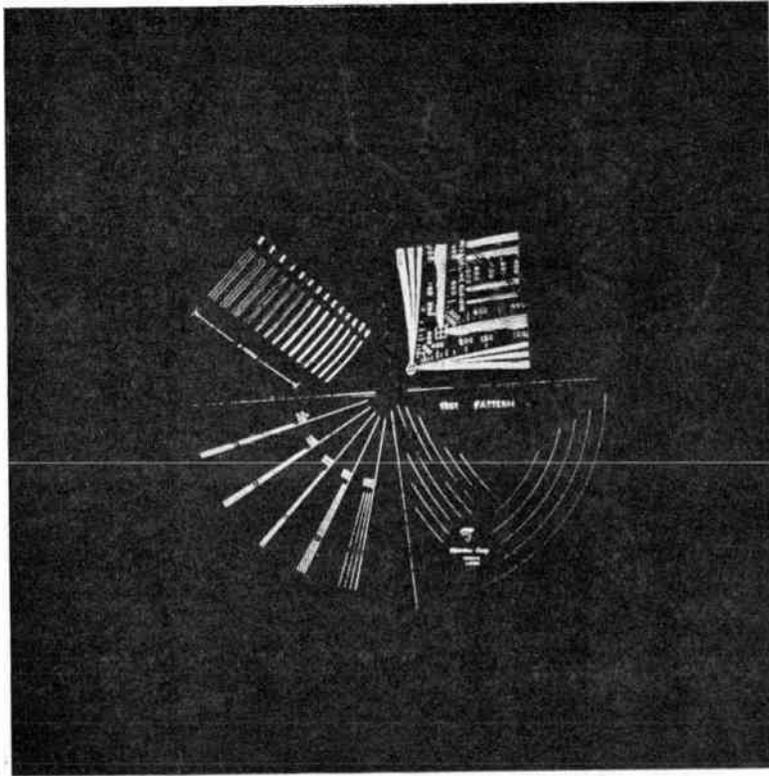


Fig. 7. Resolution pattern, 150 μ s sweep.

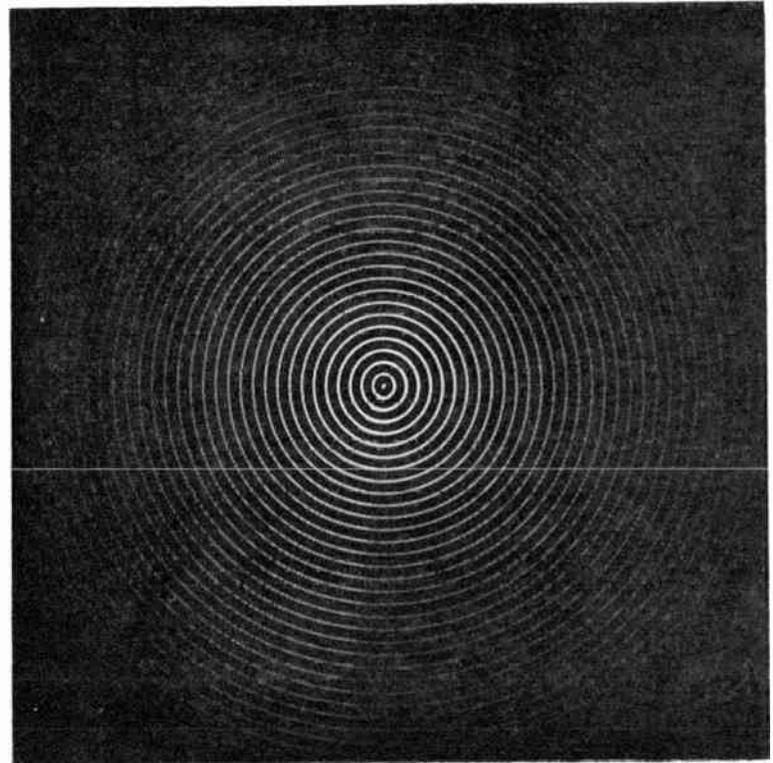


Fig. 8. 60-mile, dual range mark, 2 sweep.

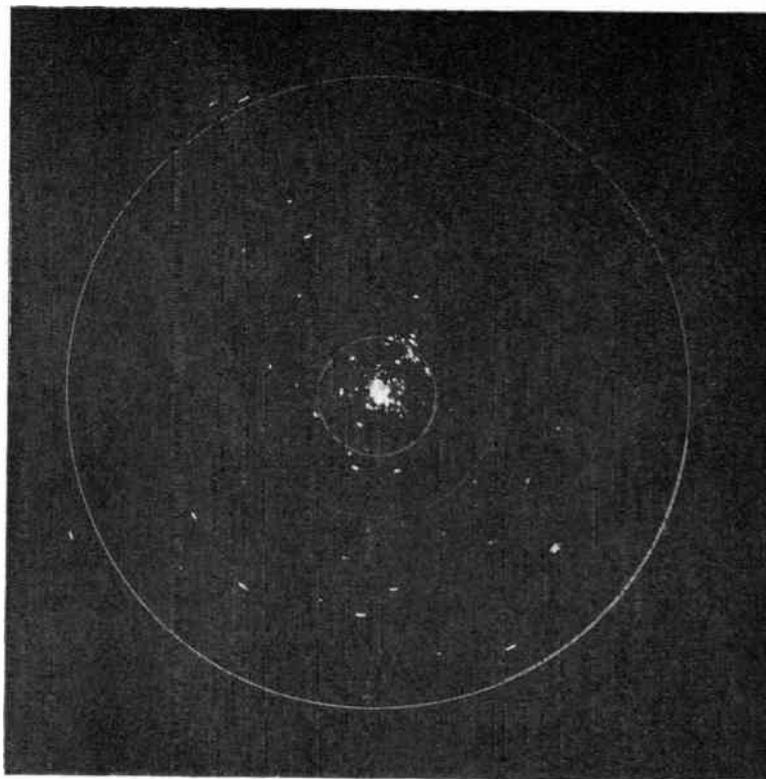


Fig. 10. 60-mile, m.t.i. video.

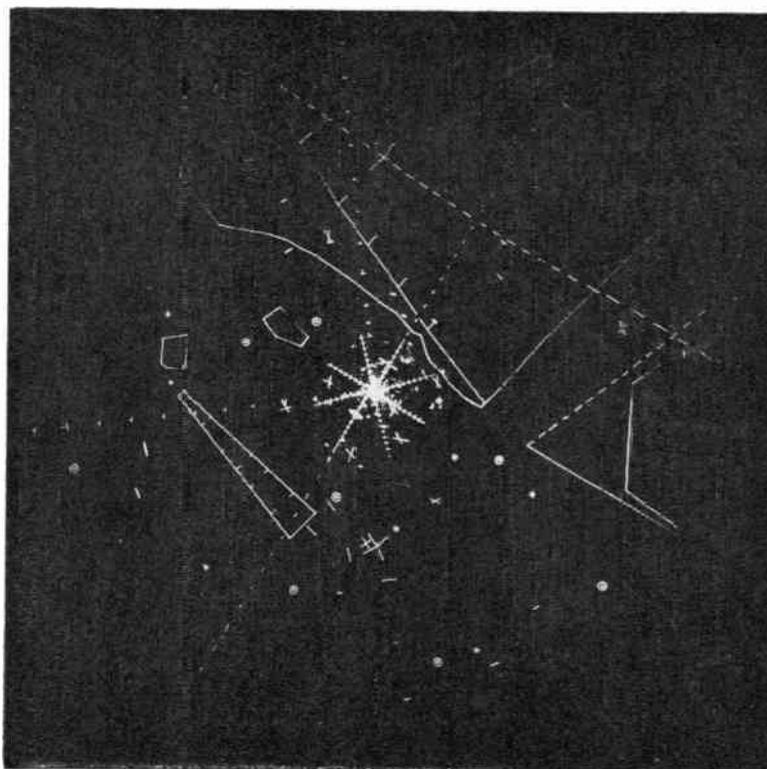


Fig. 9. 60-mile, m.t.i. video, map.

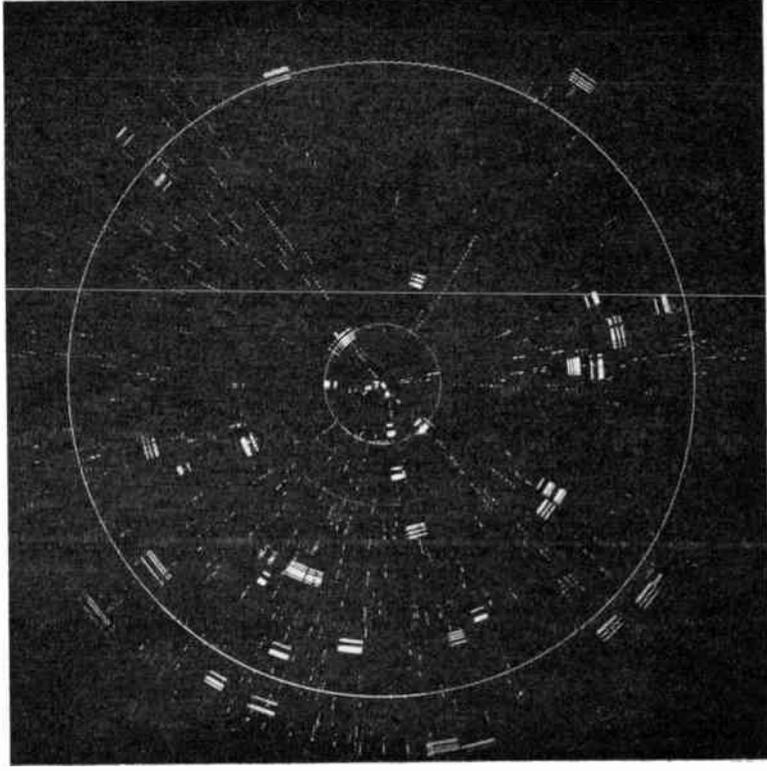


Fig. 12. Beacon video, 60-mile range.

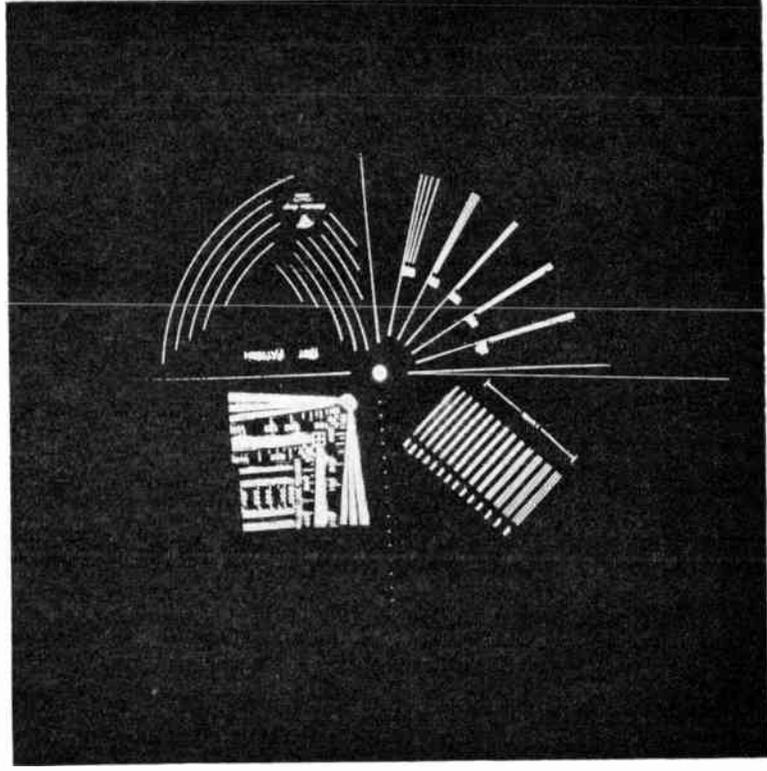


Fig. 11. Resolution map, 740 μ s sweep.

2.1. System Parameters

Tests conducted using the c.r.t.-vidicon approach to sweep-time reduction yielded data that are applicable both to analogue and digital converter techniques. A discussion of these tests follows. (See Appendices for the data.)

2.1.1. C.r.t.-phosphor-vidicon-photoconductor match

The vidicon, an 8507 separate-mesh type, was found to have a photoconductor sensitivity that was essentially flat from 5000 angstroms to 5800 angstroms (500–580 nm). The 5300 Å output of the P31 phosphor results in both proper frequency and a high efficiency. The effect of mismatch would be a loss of contrast on the radar indicator. A long persistency phosphor such as a P28 is unsatisfactory due to the trail that develops in successive sweeps. While limited storage is a desirable feature for integration of the signal, it must be held below a value that does not increase excessively azimuth target size. A discussion of storage is given in Section 2.1.5.

2.1.2. Sweep-time linearity

System linearity is affected by the camera and input p.p.i. deflection systems. In the 'breadboard' model, linearity correction was developed in the vidicon camera to compensate for non-linearities found both in the camera and the p.p.i. sweeps.

The tests consisted of sweep-reducing 10-mile range marks by a 5 : 1 ratio. The counted-down 2-mile range marks were compared on an A-scope with the compressed 10-mile marks. Coincidence between compressed 10-mile markers and non-compressed 2-mile markers was maintained to within 0.2 μs or approximately 0.1%.

The high linearity achieved is largely due to the use of single axis deflection. The large area of the vidicon that is unused invites possible use to increase system resolution. The video could perhaps be written using several Y-axis sweeps displaced only slightly in the X-axis. Since resolution is not a limiting attribute of the system, no experimentation was conducted to improve it by using more of the vidicon area.

2.1.3. Shading

Shading in the c.r.t.-vidicon reduction system is caused by misregistration of the write-read beams. Failure of the reading vidicon beam to proceed along the path of the written line results in a decreased video amplitude at the range of misregistration.

The laboratory experience has shown that misalignment can be caused by either a mechanical

movement or an electrical change. The mechanical displacement is not a problem of concern. The electrically-caused movement is produced by long-term drift in the focus of the vidicon and must be considered a system problem for solutions employing magnetic focus to achieve small spot-size.

The vidicon focus is both magnetic and electrostatic. The magnetic focus component with a change in focus coil current shifts the beam position as a result of the normal force applied to the beam. Because of this problem, electrostatic focus is preferred. The design of such a tube that will provide the required spot-size with electrostatic focus and deflection has been explored. The tube would be a single-gun storage tube of cost consistent with design goals.

2.1.4. Sweep-time ratios

The selection of a 5 : 1 sweep-time reduction ratio was based upon existing p.p.i. sweep speeds and upon the resolution requirements of the indicator display. Other benefits were also accumulated by this selection and will be discussed.

The minimum sweep-time provided by the radar console is for 6-mile range or approximately 75 μs. The offsetting at 60 miles range is not to be permitted. Offsetting can be obtained from ranges of 6 miles to 55 miles. If in the sweep-time reduction system a radius offset is made while on the 6-mile range and a constant sweep-time for all ranges of 150 μs is maintained then the maximum sweep speed of 75 μs per radius is not exceeded. This condition does not burden the deflection amplifier nor the video amplifier of the radar console. A compromise is made between sweep speed and the time available to write alpha-numerics. For a 50% reduction in sweep speed that is achieved by using 5 : 1 reduction rather than 10 : 1, a price of 10% is paid in the time made available to write characters.

2.1.5. Minimum discernible signal strength

Tests were conducted to determine the effect of sweep-time reduction of the low-level targets. Since the vidicon was known to store from sweep to sweep, an improvement in the minimum discernible signal could be expected. An earlier work¹ on signal integration of storage tubes indicated that a gain of significance could be realized. The gain in the system was substantiated by tests made.

A pulse generator was adjusted to produce simulated targets of a 1 μs wide pulse delayed out to one-half the distance between centre and edge. The noise level environment was controlled by amplifying low level radar noise to the desired amplitude. The

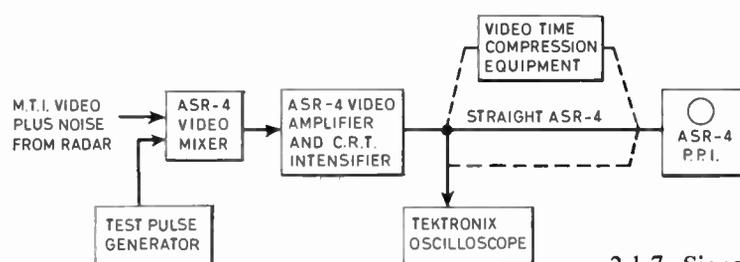


Fig. 13. Block diagram of ASR-4 sweep-time reduced video Signal/Noise ratio tests.

simulated radar target pulse and noise were connected into the sweep-time reduction system and viewed on the radar display console (see Fig. 13). For a fixed noise level, the simulated pulse was decreased in amplitude until the signal could no longer be detected. The test was repeated with the signal and noise being connected directly into the p.p.i. display console. The signal was decreased until it could no longer be detected and at this point its amplitude on the p.p.i. c.r.t. grid was measured and recorded.

Results obtained above were substantiated by a second test. Separate transfer curves through the sweep-time reduction system were derived for both radar noise and simulated target levels. The noise and target levels were compared to determine the signal enhancement from the system storage. It was found that the system integrated the target signal to twice the level of random noise. Thus for target and noise levels into the reduction system, the output target level will be twice the amplitude of the noise measured at the reduction system output.

The test results indicate that a signal can be more readily detected through the sweep-time reduction system rather than a signal fed directly into the output display. The signal enhancement is approximately 6 dB of voltage gain when operating on the linear portion of the transfer curve.

2.1.6. Radar console p.p.i. c.r.t.

The c.r.t. K1709P of the p.p.i. has a control grid range of 75 V. A beam aperture is used to achieve a spot size of approximately 0.38 mm (0.015 in) with electrostatic focus. Beam current limiting occurs non-linearly, therefore, with applied signal voltage. When sweep speed is increased to 10 times normal, the luminance of the c.r.t. fall off by approximately 3 : 1.² While this fall-off is noticeable in brightness, the comparable fall-off in brightness at a 5 : 1 sweep reduction ratio is not noticeable. A ramp voltage was applied to the input c.r.t. to provide partial compensation for high brightness contrast on the output radar indicator. Neither range nor video compensation circuits are present in the radar p.p.i. display used in the experimentation.

2.1.7. Signal/noise ratio

Signal/noise ratio measurements made using the P31 phosphor on the 5-inch c.r.t. compared peak signal output with r.m.s. camera noise. The P31 phosphor has sufficient luminescence to drive the vidicon into saturation thereby insuring a peak signal output. The signal/noise ratio determined under these conditions was found to be 39 : 1.

2.2. Resolution

Discussed under this heading are several sub-topics which relate to the resolution of the system. Among these subjects are minimum resolvable pulse width, video amplifier bandwidth, resolution of p.p.i. c.r.t., and spot-size growth.

2.2.1. Minimum resolvable pulse width

The tests conducted to determine minimum resolvable pulse width reflect the spot size limitation of the sweep-time reduction system. A constant amplitude pulse of variable width was adjusted in discrete steps through the c.r.t.-vidicon chain having a 5 : 1 sweep-time reduction ratio.

Data for this test are contained in Appendix 5.8. The data show an inability to achieve a similar ratio of reduction in video pulse width. From a maximum reduction of 3.6 : 1, the ratio diminishes to a 1 : 1 ratio occurring at an input pulse width of 0.4 μ s. Below this pulse width the sweep-time c.r.t.-vidicon stretches the video pulses.

2.2.2. Video amplifier bandwidth

As stated in Appendix 5.1, the input video amplifier has a bandwidth of less than 2.5 MHz for the terminal radar considered. Having reduced the sweep-time by 5 the bandwidth must be established for the new video amplifier. The bandwidth may be shown as:

$$b_2 = \frac{T_1}{T_2} b_1$$

where b_1, b_2 = bandwidths of input and output video amplifiers respectively,

T_1, T_2 = sweep-times of input and output video respectively.

The bandwidth of the output amplifier while being proportional has not been found in the laboratory to be directly proportional. This is due to the limited

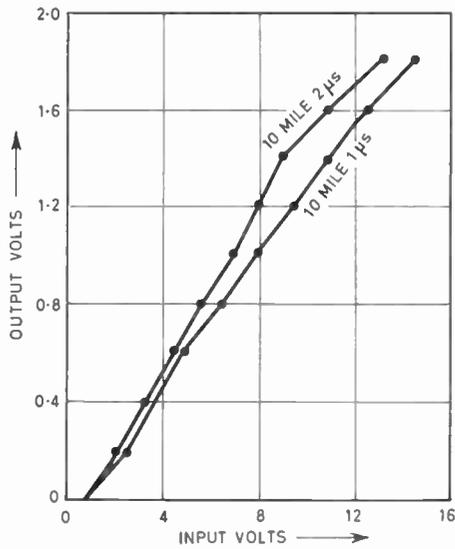


Fig. 14. C.r.t.-vidicon system transfer curve—10 mile.

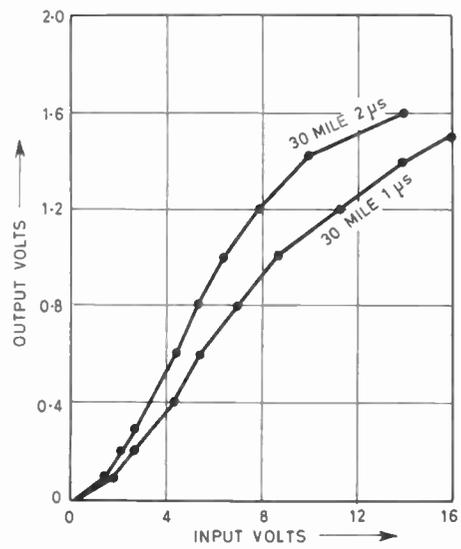


Fig. 15. C.r.t.-vidicon system transfer curve—30 mile.

band-pass of the p.p.i. c.r.t. With this in mind, the system bandwidth can be developed from:

$$1/b_s^2 = 1/b_2^2 + 1/b_T^2$$

where b_s = band-pass of the system,

b_T = band-pass of the p.p.i. c.r.t. = d/v , where the c.r.t. spot size d is 0.38 mm (0.015 in) and the spot velocity v is 9.33 in/0.1 ms.

Laboratory results have been satisfactory under the condition that b_2 is raised to 10 to 12 MHz when the reduction ratio was operated at 5 : 1 for the c.r.t.-vidicon chain.

2.2.3. Range ring resolution

One-microsecond pulses spaced 1.5 μ s from trailing-edge to leading-edge were used to generate 300 range rings on the 60-mile range scale of the radar display while the radar video was processed through the c.r.t.-vidicon chain. The range rings were spaced so that the pulses interfaced at the 50% voltage points. This resolution more than meets the specification requirements of the p.p.i. display that is satisfied by only 150 range rings. The specification requirements are described under distance resolution in Appendix 5.1.

2.2.4. Spot growth effect

The limit of the system band-pass at the indicator c.r.t. is as for most radar systems.³ When the bandwidth of the video system is given by

$$b = \frac{1}{T}$$

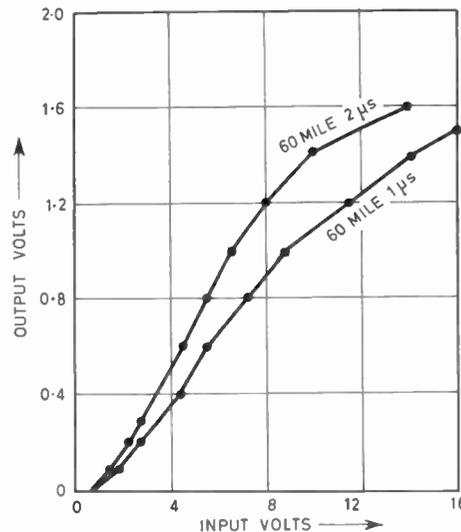


Fig. 16. C.r.t.-vidicon system transfer curve—60 mile.

and

$$\frac{d}{v} > T$$

then several periods of the video amplifier will occur during the one spot size. For the radar system detailed in the Appendix, it is approximately 4 : 1. The effect averages the signal changes occurring as the beam moves one spot size. A reduced bandwidth produces the same effect.

If a 0.5 μ s pulse is applied to the p.p.i. c.r.t., spot growth appears as 0.127 mm (0.005 in) additional spot width added to a nominal 0.38 mm (0.015 in)

spot size. When this is referred to range rings of 0.5 μs pulse width, a maximum of 300 is calculated resolvable (see Fig. 18).

In the c.r.t.-vidicon sweep-time reduction system a reduction of 30% of spot size at the p.p.i.-c.r.t. to 0.254 mm (0.010 in) will result in an increase of range ring resolution to 340 while a 30% reduction in spot size at the 5-inch c.r.t. would increase expected range ring resolution from 271 to 288. Such a reduction in the 5-inch c.r.t. would be very expensive and less effective than a corresponding reduction in the p.p.i.-c.r.t. spot size.

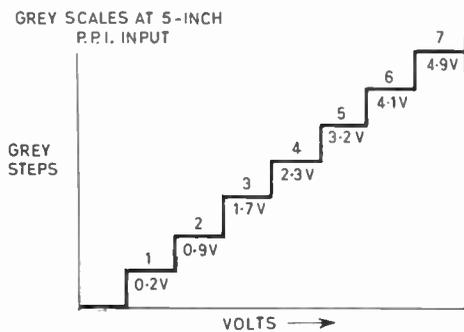
2.3. Transfer Characteristics

The transfer characteristics were derived for three sweeps, as shown in Figs. 14, 15 and 16, respectively, for 10 miles, 30 miles and 60 miles. Two pulse widths were examined for each writing speed, 1 μs and 2 μs.

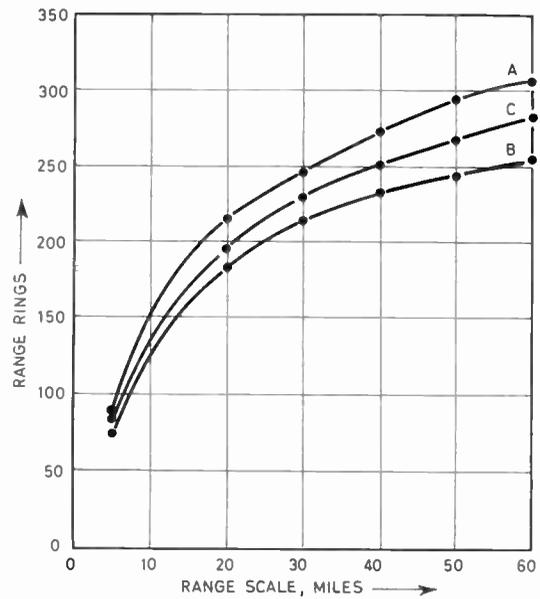
The input signal was measured at the 5-inch p.p.i. c.r.t. grid while the output signal was measured at the vidicon amplifier output. The results indicated sufficiently that the transfer function between 5-inch c.r.t. and vidicon does not vary with radar writing speed and is a straight linear line from the smallest input signal to saturation. The range of input signals vary from 0.15 to 14 V. The 0.15 V signal is limited by the inherent camera noise.

2.3.1. Grey scale

Grey scale measurements through the vidicon time reduction system were completed using a seven-step grey scale generator. The seven grey steps were applied to the 5-inch input c.r.t. grid and compared on a dual trace A-scope. A linear distribution of grey scales was achieved without difficulty. The representative data are shown in Fig. 17.



(a) Grey scales at 5 in p.p.i. input.



(b) Time compressed outputs.

Fig. 17. Grey scale for c.r.t.-vidicon.

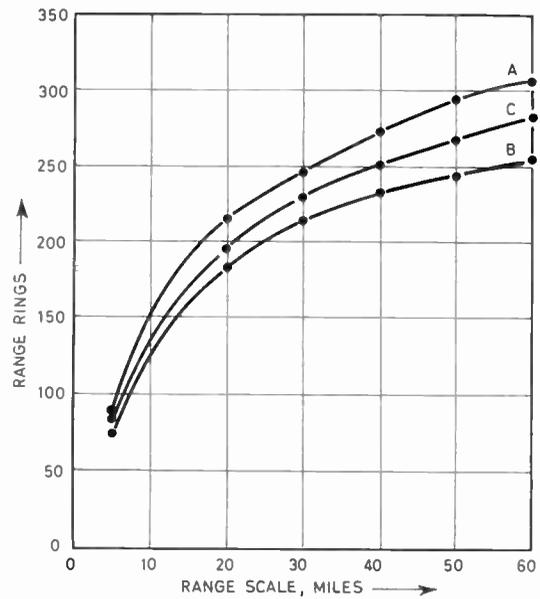


Fig. 18. Calculated range ring reduction.

Curve A: p.p.i. c.r.t., direct.

B: p.p.i. c.r.t., c.r.t.-vidicon, 5 : 1.

C: p.p.i. c.r.t., storage tube, 10 : 1.

2.3.2. Off-centring

When off-centring is activated, the range of the stored video on the write side of storage is increased as is the deflection voltage on the p.p.i. display. How these changes are accomplished is discussed below.

2.3.3. 'Write' side range change

The deflection voltage reached and the swept distance are constant for the 'write' side storage. Time is adjusted as a function of the p.p.i. range setting and the off-centring controls. The off-centring control acts

to increase range with increasing off-centring. Controls for off-centring are two independently operated linear potentiometers. These potentiometers must now control the attenuation of deflection voltage ramp to the amplifier of the 'write' side store. Maximum attenuation occurs at maximum offset. A voltage change proportional to $\sqrt{(X^2 + Y^2)}$ is required to increase the time of the video gate. The video gate controls the period of writing time for video. Both video gate termination and maximum sweep voltage should occur at the same time.

2.3.4. P.p.i. deflection voltage changes

The read-time of the storage device is assumed fixed at 150 μ s. The deflection voltage maximum must change as a function of off-centring. Again, as with the 'write' side, the off-centre output to the p.p.i. comes from two linear potentiometers. The offset voltages must be applied to adjust attenuation of the deflection voltage to the appropriate amplifier. Maximum attenuation will be applied at no offset. As offset is increased, attenuation will be decreased, maximum deflection voltage increased while sweep time held constant. The change in deflection voltage should increase linearly. The resultant deflection fields will produce the required cross-product of $\sqrt{(X^2 + Y^2)}$. A possible solution is proposed. In lieu of two potentiometers for N-S and E-W off-centre there could be a potentiometer for $\sin \theta \cos \theta$ and a second for Z. Under this condition then, $Z = \sqrt{(X^2 + Y^2)}$ would provide a direct linear correction to the 'write' side range and to the p.p.i. attenuation charge of $Z \sin \theta$ and $Z \cos \theta$ for inputs

to the E-W and N-S deflection amplifiers, respectively. On the p.p.i. the radar origin at some off-centre Z will be rotated about the p.p.i. c.r.t. centre.

3. Digital Conversion

The digital conversion technique is now a feasible one and a block diagram of a proposed design is shown in Fig. 19.

The digital conversion approach is important because it is considered possible through its use to develop more efficient methods of target detection and at the same time achieve reductions in target data transmission rates. As a minimum, these digital signals could be presented to the controller as a cross check on his present analogue-to-digital target conversion.

It is recognized that alpha-numeric must track with the target and prime radar targets require elaborate and costly detection and tracking devices. In the balance of cost of development and benefit, both operational and technical, it is hoped that the digital conversion described will tip the balance toward simplification of this detection process.

The digital sweep-time reduction technique accepts radar video in analogue form for sampling. The sampling rate, if high enough to reproduce the initial wave shape, must be at least twice the highest video frequency to be reproduced. The radar target signals of the system considered are normally of 1 to 2 μ s duration. The presence of these targets will be detected or sampled if the rate is set at 2 MHz. This sampling frequency will not aid in the discrimination of signal-

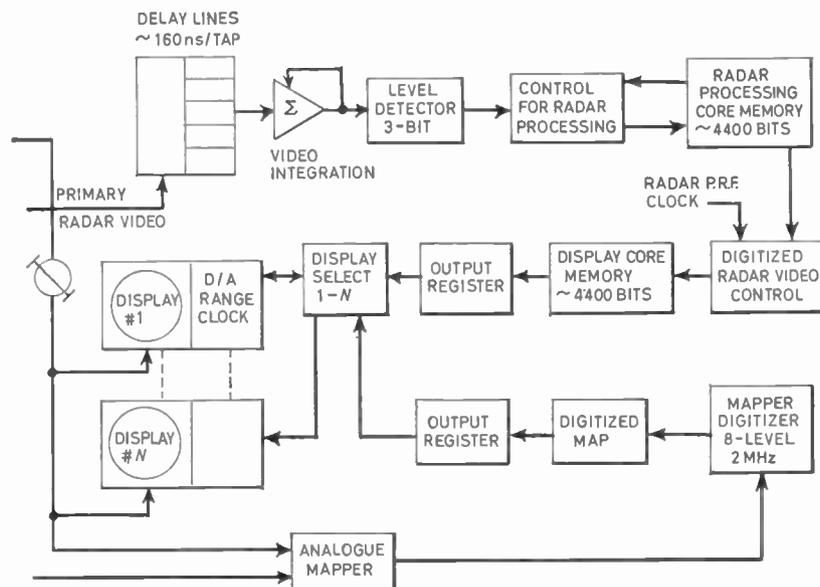


Fig. 19. Digital video sweep-time reduction block diagram.

to-noise energy pulses. When several samples of 5 MHz rate are integrated, the presence of a high energy signal such as a target results in a high integration output. The output level of the integration will therefore be weighted toward wider duration pulses. The shorter, high voltage noise spike will produce a significant output only when occurring at a high frequency. A block diagram of a digital system using such a front-end is shown in Fig. 19.

A second important feature of the digitizer is the sweep-to-sweep signal integrator. Harrington and Rogers have shown that an integration factor of 0.5 is suitable for this application.¹ Consequently, to produce the integration equivalent to that inherent for the analogue system of the vidicon or storage tube storage after read down, a digital shift-right one bit and add is incorporated. The product of the digitizer is a sampled video described in three bits sampled one each microsecond. The data store must be accessed at a rate dependent upon compression ratio desired and digital store word size. For a ratio of 5 : 1 and a word size of 15 bits the word output rate is computed to be 1 MHz. To accomplish the output, dual registers are shown in Fig. 19. The radar video changing at 0.17 μ s rate is summed in an operational amplifier whose output is sampled at a 2 MHz rate and coded by a three-bit byte to one of eight levels. The three-bit bytes are summed with the corresponding three bits from core memory to provide sweep to sweep integration. The control for radar processing has in part two adders overlapped in time to allow the radar processing and input/output of the radar processing memory to be matched. The core memory has a 1 μ s cycle-time for a 15-bit word. During radar video dead-time the contents of the radar processing memory, 4400 bits, are transferred to the display core memory resulting in the updating of all information in the display memory. The display memory is then ready to service the displays. Since the display time is a constant, the clock-out rate to each display has to be a variable. Each display is set to the correct clock rate for its range and the information fed to the appropriate d-a converter. The time period being fixed, the information rate has to change in order to provide the proper time spacing for the 60-mile range and with range change an appropriate change in clock rate would be required. If the time for a display is made 100 μ s then the clock must move 1466 bytes of three bits each byte in the 100 μ s period for the 60-mile range or a clock requirement for 14.6 MHz. A better choice might be 150 μ s and a 10 MHz (approximately) clock. The clock for the six-mile range would be 1.46 MHz or 1 MHz. The number of separate ranges that is to be supplied by one display memory is fixed by the choice of this sweep-time. As many display memories as desired can be added in parallel to the radar processing core memory.

4. References

1. J. V. Harrington and T. F. Rogers, 'Signal-to-noise improvement through integration in a storage tube', *Proc. Inst. Radio Engrs*, 38, pp. 1197-1203, October 1950.
2. 'K1709P7 Study', Unpublished Report, National Aviation Facilities Experimental Center, October 1967.
3. L. N. Ridenour (Ed.), 'Radar System Engineering', p. 44, M.I.T. Radiation Laboratory Series, Vol. 1 (McGraw-Hill, New York, 1947).

5. Appendix

5.1. Radar and P.P.I. Console Specifications

Range	Minimum 6 miles Maximum 60 miles
Pulse repetition frequency	1200/s (nominal)
Pulse width	0.8 μ s
Video amplifier bandwidth	2.5 MHz
Antenna beam width	1.5°
P.P.I. display c.r.t.	15 in, K1709P7
Deflection	Push-pull, resolved sweep, magnetic
Focus	Electrostatic
Decentering	2 radii up to 20-mile range
Sweep linearity	$\pm 1\%$ of c.r.t. radius at cardinal azimuth, $\pm 2\%$ at other azimuth angles
Sweep-trace line width	Maximum 0.02 inches
Maximum allowable spot growth ratio	1.5
Distance resolution (the greater of)	1% of sweep range or 1½ times pulse width
Sweep dead time, 60-mile range	91 μ s
60-mile range sweep time	741 μ s

5.2. P.P.I. C.R.T. Deflection-yoke Data

Horizontal deflection inductance	8.3 mH
Vertical deflection inductance	9.3 mH
Horizontal deflection resistance	22 Ω
Vertical deflection resistance	22 Ω
Critical damping resistance	2.7 k Ω
Resonant frequency	75 to 80 kHz

5.3. Deflection Amplifier

A sine-wave signal generator output of 0.5 peak-to-peak was applied to the input of the p.p.i. horizontal deflection amplifier. Data for the test are as follows:

Signal generator frequency kHz	Deflection amplifier cathode voltage across half 51.1Ω (peak-to-peak)	Horizontal deflection yoke $\times 11$ (peak-to-peak)
50	2.8	38.0
100	2.4	20.0
150	—	9.0
200	2.0	4.4
300	1.6	3.2
400	2.4	
500	2.4	
600	2.8	
700	2.4	
1030	2.4	

5.4. Deflection System Settling Time

Settling time of the horizontal deflection system of the p.p.i. console was measured and determined to be approximately 34 μ s.

Measurement was made by operating the p.p.i. on the 6-mile range with the azimuth drive disabled.

A pulse with a fall-time of less than 1 μ s, and of sufficient amplitude to drive the sweep 1 radius was applied to the horizontal deflection amplifier. A dual pulse generator of adjustable pulse spacing was used to measure the fall-time (T_1 to T_2) as 9 μ s and time to settling T_2 to T_3 as 24.6 μ s.

5.5. Video Amplifier Response

Input sine-wave 0.5 V p-p kHz	Output V (p-p)
50	10
100	10
1000	9
1760	6
2500	4
3000	3
4000	2
5000	1.5
6000	1.0

5.6. Minimum Discernible Signal for C.R.T.-Vidicon Sweep Reduction

Radar noise average volts	Simulated target pulse-width μ s	Pulse amplitude input peak volts	Non-reduced peak volts	Reduced peak volts
10	2	4.8	1.92	2.5
10	2	3.5	1.35	2.6
10	2	3.2	1.28	2.5
10	2	4.0	1.67	2.4
5	2	1.7	1.06	2.6
5	2	1.9	1.36	1.4
5	2	2.4	1.71	1.4
Re-aligned				
10	2	3.7	1.42	2.6
10	2	3.2	1.41	2.2
10	2	3.4	1.41	2.4
10	1	3.2	1.07	3.0
10	1	3.8	1.36	2.8
10	1	3.5	1.34	2.6

5.7. Substantive Data for Minimum Discernible Signal

Input signal level† (V) [A]	Radar noise [B]	Output level (V) Range ring pulses 2 μ s [C]	Trig. rate pulses 2 μ s [D]	Gain (dB) $20 \log \frac{[C]}{[B]}$	Gain (dB) $20 \log \frac{[D]}{[B]}$
2	0.16	0.2	0.24	1.58	3.5
4	0.20	0.3	0.40	3.50	6.02
8	0.30	0.55	0.65	5.22	5.30
12	0.40	0.65	0.80	4.20	6.02
16	0.45	0.70	0.90	3.81	6.02

† Average volts as recorded on A-scope. Time compression = 5 : 1. Sweep 150 μ s. Signal gain through t.-c. system.

5.8. Spot Growth Calculations

Spot growth is calculated through the chain of storage tube sweep-time reduction system is calculated. 5-inch high-resolution c.r.t., vidicon, and finally The first system calculations are based on a reduction p.p.i. c.r.t. In addition, spot growth for a single-gun ratio of 5 : 1 while the second is a ratio of 10 : 1.

(a) 5-in c.r.t. and p.p.i. c.r.t.

Range (miles) [A]	Range time (μ s) [B]	Motion/range mark time [C] $R[B] \times 0.5 \times 10^{-6}$ in $\times 10^{-3}$		Range mark size [D] Spot size + [B] in $\times 10^{-3}$	
		5-in c.r.t.	p.p.i. c.r.t.	5-in c.r.t.	p.p.i. c.r.t.
60	740	1.86	4.87	3.36	19.87
50	617	2.23	5.83	3.73	20.83
40	494	2.79	7.29	4.29	22.29
30	371	3.71	9.70	5.21	24.70
20	247	5.57	13.13	7.07	28.13
10	124	11.1	29.0	12.6	44.0
5	62	22.2	58.1	23.7	73.1

5-in c.r.t. spot size 0.0015 in (0.0381 mm) Useful radius (R) of p.p.i. 7.2 in
P.p.i.-c.r.t. spot size 0.015 in (0.381 mm) Used writing length of c.r.t. 2.75 in
Range mark pulse time 0.5

Calculated range rings possible on p.p.i. c.r.t. is given by $\frac{1.7 \times 7.2}{2 \times [D]}$

(b) Vidicon

Range	Vidicon image size [F] $\frac{0.5}{2.75} \times [D]$ in $\times 10^{-3}$	Image length read-time [G] $[F] \times \frac{150}{0.5} \times 10^{-9}$ μ s	P.p.i. image motion [H] $\frac{7.2 \times [G]}{150 \times 10^{-6}}$ in $\times 10^{-3}$	P.p.i. image size [J] [H] + 15 in $\times 10^{-3}$
60	0.612	0.184	8.8	23.8
50	0.679	0.204	9.8	24.8
40	0.781	0.234	11.2	26.2
30	0.947	0.284	13.6	28.6
20	1.29	0.386	18.5	33.5
10	2.29	0.687	33.0	48.0
5	4.31	1.293	62.0	77.0

Vidicon image length 0.5 in Sweep-time 150 μ s
Vidicon resolution 800 linear/dib Spot size 0.00031 in
Image growth through lens assumed nil.

Calculated range rings possible is given by $\frac{1.7 \times 7.2}{2 \times [J]}$

(c) Single Gun Storage Tube

Range	Motion/range mark [K]	Image size [L]	Image length read-time [M]	P.p.i. image movement [N]	Image size at p.p.i. [P]
	$\frac{4.5 \times 0.5}{[B]}$	[K] + 1	$[L] \times \frac{75}{4.5} \times 10^{-9}$	$\frac{7.2}{75 \times 10^{-6}} \times [M]$	15 + [N]
miles	in $\times 10^{-3}$	in $\times 10^{-3}$	μs	in $\times 10^{-3}$	in $\times 10^{-3}$
60	3.04	4.04	0.0675	6.48	21.48
50	3.64	4.64	0.0773	7.42	22.42
40	4.56	5.56	0.0927	8.91	23.91
30	6.07	7.07	0.118	11.3	26.3
20	9.10	10.10	0.168	16.1	31.1
10	18.1	19.1	0.318	30.3	45.3
5	36.3	37.3	0.621	59.6	74.6

Read sweep-time 75 μs Sweep length 4.5 in Target diameter 5 in nominal
 Spot size 0.001 in (0.025 mm) Tube type FW-202

Calculated number of range rings possible are $\frac{1.7 \times 7.2}{2 \times [P]}$

5.9. Calculated Range Rings

Range scale	P.p.i. c.r.t. direct (No sweep-time reduction)	Range rings	
		5 : 1 sweep-time reduction via c.r.t.-vidicon	10 : 1 sweep-time reduction via storage tube
		$\frac{1.7 \dagger \times 7.2}{2 \times [D]}$	$\frac{1.7 \times 7.2}{2 \times [J]}$
60	308	256	284
50	293	246	273
40	274	233	255
30	247	214	242
20	217	183	197
10	139	127	135
5	84	70	82

† Range rings factor assumed per television line pair.

Manuscript first received by the Institution on 1st June 1967 and in revised form on 1st January 1968.
 (Paper No. 1198/AMMS14.)

© The Institution of Electronic and Radio Engineers, 1968

Synthesis of the Ungrounded Inductor with Grounded Amplifier Realizations

By

A. W. KEEN,
M.Sc., C.Eng., F.I.E.E.†

AND

Mrs. J. L. GLOVER, B.Sc.†

Summary: The ungrounded inductor is approximated closely by the transmission matrix $(1, Ls; g, 1)$, where $L = CR^2$, by making g sufficiently small. A representation of this matrix is obtained by a form of synthesis procedure and realized in active-RC form, with L derived by inversion of a capacitor C , using a minimal number of grounded operational amplifiers of only one type. The theory has been confirmed experimentally by replacing an actual coil in single sections of prototype and shunt m -derived low-pass filter.

List of Symbols

$A, B; C, D$ $a, b; c, d$ }	elements of the transmission matrix
g	a small conductance
K	numerical constant
$[L]_{Nr}$	non-reciprocal ungrounded inductance matrix
$[L]_{nr}$	as $[L]_{Nr}$, with $N = n, C = g$.
$[L]$,	reciprocal ungrounded inductance matrix
m	m -factor of m -derived filter
N	common value of A and D in the non-reciprocal factor matrix $(N, 0; 0, N)$
R_{ij}	real part of the impedance matrix element Z_{ij} where $i = 1, 2; j = 1, 2$
s	complex-frequency variable
$[T], [t]$	transmission matrices
$[Y]$	short-circuit admittance matrix
$[Z]$	open-circuit impedance matrix
Z_{in}, Z'_{in}	input impedance at port 1
Z_L	load impedance at port 2
Z_{XY}	arbitrary impedance across (X, Y)
$\alpha, \beta; \gamma, \delta$	elements of the transmission matrix
$\Delta, \Delta T$	determinant of the transmission matrix

1. Introduction

A number of recent publications¹⁻⁵ have considered the problem of simulating an ungrounded inductor, which arises in the active-RC realization of low-pass and band-pass filters. Three types of realization have been proposed: in one type a grounded shunt capacitor is cascaded between a pair of actively-realized gyrators;¹ in another type a single ungrounded gyrator is loaded with a capacitor;^{2, 3} in the third

two grounded inductors are combined into a push-pull form.^{4, 5} The first and third types require two gyrators (or their equivalents) and, when realized with standard amplifier units, require a large number of units, in some cases of more than one kind; the second type requires ungrounded supplies for the active elements of the gyrator, or a special supply unit.² In no case has there been any derivation of the proposed realization from a mathematical specification of the required inductor. As an extension of an earlier publication,⁶ in this paper a synthesis approach to the problem is adopted and a form of realization, with several variants, that employs a minimal number of standard (differential-input, grounded output operational amplifier) active units, all of the same kind, is obtained. In order to clarify the formal procedure an elementary intuitive approach is first described.

2. An Intuitive Approach

Consider a symmetrical non-reciprocal network having $A = D = N$ (possibly complex), and with $B = C = 0$. Let an ideal ungrounded inductor, as defined by its transmission matrix,

$$L_r = \begin{bmatrix} 1 & Ls \\ 0 & 1 \end{bmatrix} \quad \dots\dots(1)$$

be cascaded with the network (Fig. 1). This arrangement gives a non-reciprocal form of inductor, which is characterized by the overall matrix:

$$\begin{aligned} L_{Nr} &= \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix} \begin{bmatrix} 1 & Ls \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & Ls \\ 0 & 1 \end{bmatrix} \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix} = \begin{bmatrix} N & NLs \\ 0 & N \end{bmatrix} \quad \dots\dots(2) \end{aligned}$$

The driving-point impedance at either port, with a load Z_L on the other, will be

$$Z_{in} = (AZ_L + B)/(CZ_L + D) = \frac{NZ_L + NLs}{N} = Z_L + Ls \quad \dots\dots(3)$$

which is independent of the order in which the inductor matrix and the non-reciprocal factor are multiplied

† School of Electrical Engineering, Bath University of Technology, Bristol.

and also of the value of N , i.e. the driving-point properties of the inductor are invariant to cascade connection with the non-reciprocal N -factor matrix.

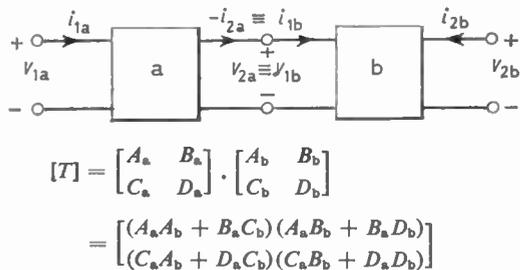


Fig. 1. Cascade form of the general two-port networks shown in Figs. 2(a) and (b), represented by the transmission matrix.

We may make the ideal inductor matrix non-reciprocal, without altering its A , B and D elements, simply by making $C > 0$; we will put $C = g$, a conductance, to obtain

$$L_{nr} = \begin{bmatrix} 1 & Ls \\ g & 1 \end{bmatrix}; \quad \Delta = 1 - gLs \quad \dots\dots(4)$$

Then we may factor out the root of the determinant $\sqrt{\Delta}$, as follows:⁷

$$\begin{bmatrix} 1 & Ls \\ g & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{\Delta} & 0 \\ 0 & \sqrt{\Delta} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\Delta}} & \frac{Ls}{\sqrt{\Delta}} \\ \frac{g}{\sqrt{\Delta}} & \frac{1}{\sqrt{\Delta}} \end{bmatrix} \quad \dots\dots(5)$$

the first factor of which is of the same type as the N factor in eqn. (2). This factor does not affect the driving-point properties of L , and the second factor is a reciprocal modification of eqn. (4). The latter will transform a load impedance in either direction as follows:

$$Z'_{in} = \frac{1Z_L + Ls}{gZ_L + 1} \quad \dots\dots(6)$$

and by making g sufficiently small we may make

$$Z'_{in} \simeq Z_{in} = Z_L + Ls \quad \dots\dots(7)$$

This is the same result as obtained before and shown in eqn. (3). However, by considering the reciprocal (admittance) expression:

$$Y'_{in} = \frac{gZ_L + 1}{Z_L + Ls} = \frac{1}{\frac{Ls}{g} + 1} + \frac{1}{Ls + Z_L} \quad \dots\dots(8)$$

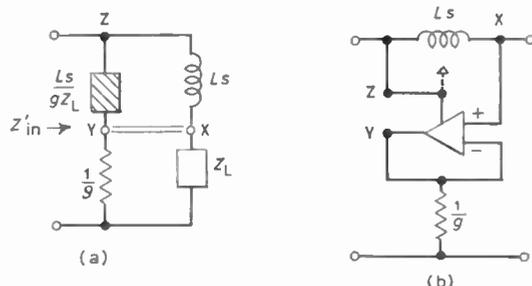
shown in Fig. 2(a), it may easily be checked that points X and Y have exactly the same potential (the configuration is a balanced bridge) and may therefore be joined by an arbitrary impedance Z_{XY} . In effect, therefore, L and L/gZ_L are in parallel so, if

Z_L is real (R_L), they will have a resultant inductance given by:

$$L' = \frac{L \frac{L}{gR_L}}{L + \frac{L}{gR_L}} = \frac{L}{1 + gR_L} \quad \dots\dots(9)$$

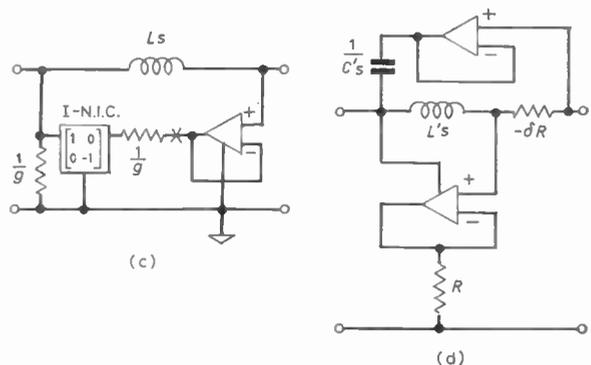
For

$$gR_L \ll 1, \quad L' \simeq L$$



(a) Equivalent circuit of the driving-point impedance (eqn. 8). (Points X and Y are at the same potential.)

(b) Replacement of the equivalent impedance branch Ls/gZ_L in (a) by a differential operational amplifier.



(c) Alternative form of (b) that allows the use of grounded amplifiers.

(d) A more general form of (b) which is equivalent to Fig. 3(a) for the case where $R_b \neq R_a$.

Fig. 2. A non-reciprocal generalization of the ungrounded inductor (eqn. 4).

A balance can be enforced⁸ by using a single differential-input ideal (operational) amplifier connected in the manner shown in Fig. 2(b). This amplifier acts as a voltage follower to make the voltage at Y follow that at X and the output current path of the amplifier simulates the equivalent impedance branch L/gZ_L , whatever the value of Z_L may be. There is a practical difficulty with the single amplifier arrangement in that the grounded side of a practical amplifier

will occur on the high side of the input terminal-pair so the inductor will no longer be ungrounded (this is an example of the way in which practical realization of inductance tends to come out with one side grounded). By using the two-amplifier arrangement shown in Fig. 2(d) this difficulty may be overcome. In either case, we now have a practical non-reciprocal active network with the following matrices:

$$T_f = \begin{bmatrix} 1 & Ls \\ g & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} \frac{1}{Ls} & g - \frac{1}{Ls} \\ -\frac{1}{Ls} & \frac{1}{Ls} \end{bmatrix}, \quad Z = \begin{bmatrix} \frac{1}{g} & \frac{1}{g} - Ls \\ \frac{1}{g} & \frac{1}{g} \end{bmatrix} \quad \dots\dots(10)$$

which have been proved to have driving-point properties similar to the coil by itself.

At first sight no advantage seems to have been gained, but we have shown the realizability of the non-reciprocal matrix represented by eqn. (4) and we now have a more general network that has a Z- as well as a Y-matrix. Also, this is an active-RC network and it is now well known that with a network of this type there is the possibility of simulating a coil. The remaining problem, therefore, is to develop the circuit further in such a way that the coil is eliminated in favour of one (or more) capacitor(s).

3. A Synthesis Procedure

We will now adopt a synthesis approach, taking the matrix of eqn. (4) as a realizable approximation to the ideal ungrounded inductor, with the value of *g* sufficiently small, and we will attempt to derive a realization from the defining matrix by expanding it into a chain of factors all of which are identifiable with a corresponding cascade of realizable units. If we are completely to avoid the need of inductors we will require at least one capacitor and we will expect this to occur in a position dual to that occupied by *L*, namely in shunt with the transmission path (but not necessarily grounded); hence the expansion may be assumed to have the form

$$[T]_{nr} = \begin{bmatrix} \alpha & \beta \\ y & \delta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Cs & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \dots\dots(11a)$$

$$= \begin{bmatrix} \alpha a + \beta(Cs) + \beta c & \alpha b + \beta(Cs) + \beta d \\ y a + \delta(Cs) + \delta c & y b + \delta(Cs) + \delta d \end{bmatrix} \quad \dots\dots(11b)$$

and our problem is to identify this with the approximating matrix (4), in which we will put

$$L = CR_1^2, \quad g = 1/R_2$$

On equating corresponding elements in eqns. (4) and (11b) the following set of equations is obtained:

$$\begin{aligned} \text{(i)} \quad & \alpha a + \beta(Cs)a + \beta c = 1 \\ \text{(ii)} \quad & \alpha b + \beta(Cs)b + \beta d = CR_1^2 s \quad \dots\dots(12) \\ \text{(iii)} \quad & ya + \delta(Cs)a + \delta c = \frac{1}{R_2} \\ \text{(iv)} \quad & yb + \delta(Cs)b + \delta d = 1 \end{aligned}$$

A further relationship which can be obtained is:

$$\frac{a}{b} = \frac{1 - \beta c}{CR_1^2 s - \beta d} = \frac{\frac{1}{R_2} - \delta c}{1 - \delta d} \quad \dots\dots(13)$$

$$\text{or} \quad 1 - \beta c - \delta d = \frac{CR_1^2 s}{R_2} - \frac{\beta d}{R_2} - \delta c CR_1^2 s$$

At some point in the development we will have to assign a resistance value (*R*₁ or *R*₂) to β or *b*, or a conductance value to *y* or *c*. For example, putting $-\beta d = R_2$, with $\beta = R_2$, and $d = -1$, eqn. (13) reduces to

$$-R_2 c + \delta = \frac{CR_1^2 s}{R_2} - \delta c CR_1^2 s$$

and this equation can be satisfied with $\delta = 0$ and $c = -CR_1^2 s/R_2$. Substituting back into eqn. (12) (for required values of *L* and *R*₁ and *R*₂) four equations in four unknowns are obtained. The remaining factor-matrix elements are as follows

$$a = 1, \quad b = R_2; \quad \alpha = 1 - CR_1 s + CR_1^2 s/R_2, \quad y = \frac{1}{R_2}$$

This complete particular evaluation of eqn. (11a) is, therefore,

$$[T]_{nr} = \begin{bmatrix} 1 - CR_1 s + CR_1^2 s/R_2 & R_2 \\ 1/R_2 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ Cs & 1 \end{bmatrix} \begin{bmatrix} 1 & R_2 \\ -CR_1^2 s/R_2 & -1 \end{bmatrix} \quad \dots\dots(14)$$

and if *R*₁ = *R*₂ = *R* (say), this reduces to

$$[T]_{nr} = \begin{bmatrix} 1 & R \\ 1/R & 0 \end{bmatrix}^{(t_1)} \begin{bmatrix} 1 & 0 \\ Cs & 1 \end{bmatrix}^{(t_2)} \begin{bmatrix} 1 & R \\ -Cs & -1 \end{bmatrix}^{(t_3)} \quad \dots\dots(15)$$

The solution of eqn. (15) is not unique. Another relationship contained in eqn. (12) is (with *R*₁ = *R*₂ = *R* again)

$$\frac{\delta}{\beta} = \frac{1 - yb}{CR^2 s - ab} = \frac{1/R - ya}{1 - \alpha a} \quad \dots\dots(16)$$

Putting $-\alpha b = R$, with $\alpha = 1$ and $b = -R$, and proceeding as before, we obtain an alternative

expansion

$$[T]_{nr} = \begin{bmatrix} 1 & -R \\ C_s & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ C_s & 1 \end{bmatrix} \begin{bmatrix} 0 & -R \\ -1/R & -1 \end{bmatrix} \dots\dots(17)$$

Using the number of zero elements of each factor in eqns. (15) and (17) as a measure of their degeneracy we may make further expansions to obtain more elementary sections. For example, the first factor in eqn. (15) may be expanded as

$$[t_1] = \begin{bmatrix} 1 & R \\ 1/R & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/R & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & R \\ 0 & 1 \end{bmatrix} \dots\dots(18)$$

which is a current-inverting type of negative imittance converter (n.i.c.) loaded by a shunt resistance on the left and by a series resistance on the right, as shown in the left-hand part of Fig. 3(a).

At each stage there is also the possibility of converting to the equivalent [Y]-matrix and expanding this as a sum, with a corresponding shunt-connected set of sub-networks. For example, reverting to eqn. (15), first multiply the second and third factors together in order to keep the capacitance in one section, to get

$$[t_2 t_3] = \begin{bmatrix} 1 & 0 \\ C_s & 1 \end{bmatrix} \begin{bmatrix} 1 & R \\ -C_s & -1 \end{bmatrix} = \begin{bmatrix} 1 & R \\ 0 & CR_s - 1 \end{bmatrix} \dots\dots(19)$$

and then convert to the equivalent [Y]-matrix

$$[Y]_{23} = \begin{bmatrix} C_s - \frac{1}{R} & -C_s + \frac{1}{R} \\ -\frac{1}{R} & \frac{1}{R} \end{bmatrix} = \begin{bmatrix} C_s & -C_s \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{R} & \frac{1}{R} \\ -\frac{1}{R} & \frac{1}{R} \end{bmatrix} \dots\dots(20)$$

The first of these admittance components can be regarded as that of an ungrounded capacitor with its second row reduced to zero, and this may be obtained by cascading the capacitor with a reversed voltage-follower stage. The second component may be identified by reverting to [T]-matrix form and expanded as follows:

$$[t_3] = \begin{bmatrix} 1 & R \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & R \\ 0 & 1 \end{bmatrix} \dots\dots(21)$$

and then represented in a way similar to eqn. (18). The complete representation, which can be realized with only one type of operational amplifier, is shown in Fig. 3(a). The required equivalent of Fig. 2(b), (c), with a capacitor C replacing the inductance L is obtained. Reverting to Fig. 2(c) it can be seen that the

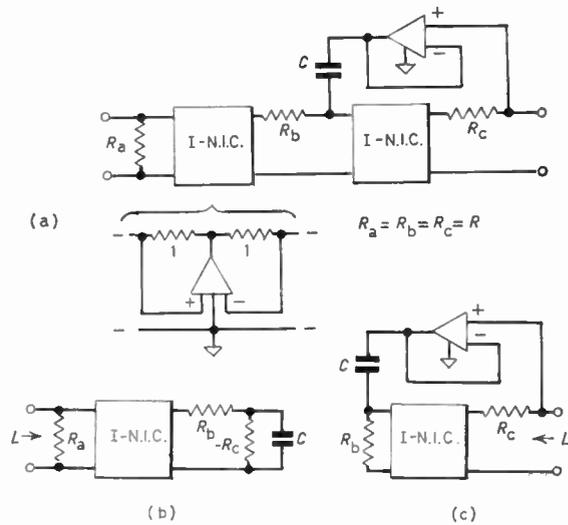


Fig. 3. (a) The synthesized equivalent of Fig. 2(c) in which the actual coil L of the latter is replaced by a capacitor C. The circuit used for the two I-n.i.c.s is shown inset. (b) Reduction of (a) when the right-hand port is short-circuited. (c) Reduction of (a) when the left-hand port is short-circuited. (N.B. (b) and (c) are both gyrator realizations.)

actual inductor L has been replaced by a capacitor between the voltage follower and its load, and from their junction there is now an additional transmission path in shunt with the follower consisting of another I-n.i.c. loaded on its right-hand side with a resistor Rc.

The stability of the non-reciprocal realization will now be considered. For this purpose it is assumed that all three of the resistors Ra, Rb and Rc have difference values. Then the transmission matrix corresponding to (4) will be

$$[T]_{nr} = \begin{bmatrix} 1 & CR_b R_c s + R_c - R_b \\ 1/R_a & \left(\frac{R_b}{R_a} - 1\right) R_c C_s + \frac{(R_c - R_b)}{R_a} + 1 \end{bmatrix} \Delta T = 1 - CR_c s \dots\dots(22)$$

The general (non-reciprocal) stability criterion for the two-port network is given by Llewellyn,⁹ but, if we factor out the non-reciprocal component, as in eqn. (5), we need only apply the simpler criterion of Gewertz^{10, 11} to the remaining reciprocal factor, namely

$$[T] = \begin{bmatrix} \frac{1}{\sqrt{\Delta T}} & \frac{CR_b R_c s + R_c - R_b}{\sqrt{\Delta T}} \\ \frac{1}{\sqrt{\Delta T} R_a} & \frac{1}{\sqrt{\Delta T}} \left\{ \left(\frac{R_b}{R_a} - 1\right) R_c C_s + \frac{(R_c - R_b)}{R_a} + 1 \right\} \end{bmatrix} \sqrt{\Delta T} = \sqrt{1 - CR_c s} \dots\dots(23)$$

constant-*k* filter and for a full-section of shunt *m*-derived filter, the circuits of which are shown in Fig. 5, are given. The one section is a particular case ($m \leq 1$) of the other and so only the theory of the more general section need be considered. The transmission was measured as a voltage-transfer ratio between fixed resistance terminations.

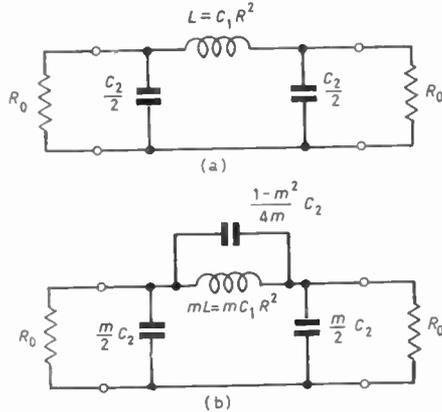


Fig. 5. Filter sections used for testing the simulated ungrounded inductor.

- (a) Prototype, $m = 1$.
- (b) Shunt *m*-derived type, $m = 0.6$.

An expression for the theoretical gain is obtained by first considering the $[T]$ -matrix of the inductance approximation (eqn. (4)) with the value modified by the *m* factor (normally $0 < m \leq 1$). This latter value is given by:

$$T_f = \begin{bmatrix} 1 & mC_1 R^2 s \\ 1/R & 1 \end{bmatrix}, \quad T_b = \frac{1}{1 - mC_1 R s} T_f \quad \dots\dots(27)$$

for the forward and backward directions respectively. Converting to *Y* form and adding the *Y*-matrix for the filter capacitor network, together with a load conductance G_0 we obtain the total *Y*-matrix. The voltage transfer ratio is given by the ratio: $-y_{22}/y_{21}$ of the elements of the total *Y*-matrix. For example, in the forward direction:

$$Y_f = \frac{1}{mC_1 R^2 s} \begin{bmatrix} 1 & mC_1 R s - 1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} \left(\frac{mC_2 s}{2} + \frac{1-m^2}{4m} C_2 s \right) & \frac{m^2-1}{4m} C_2 s \\ \frac{m^2-1}{4m} C_2 s & \left(\frac{1-m^2}{4m} C_2 s + \frac{m}{2} C_2 s + G_0 \right) \end{bmatrix} \quad \dots\dots(28)$$

$$\left. \frac{v_1}{v_2} \right|_f = - \left. \frac{y_{22}}{y_{21}} \right|_f = 1 + \frac{4mC_1 R^2 G_0 s + 2m^2 C_1 C_2 R^2 s^2}{4 + (1-m^2)C_1 C_2 R^2 s^2} \quad \dots\dots(29)$$

and in the backward direction:

$$\left. \frac{v_1}{v_2} \right|_b = 1 + \frac{4mC_1 R^2 (G + G_0) s + 2m^2 C_1 C_2 R^2 s^2}{4(1 - mC_1 R s) + (1-m^2)C_1 C_2 R^2 s^2}, \quad G \equiv \frac{1}{R} \quad \dots\dots(30)$$

For an actual coil of inductance *L* and series resistance *r*

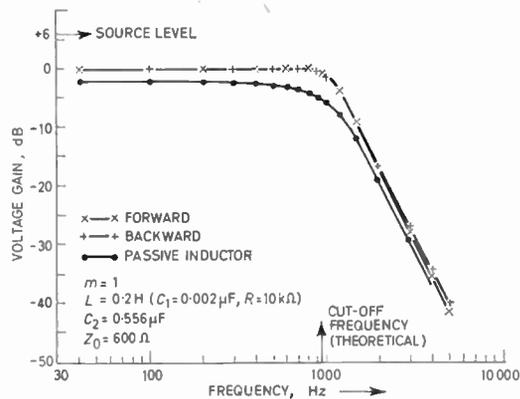
$$\left. \frac{v_1}{v_2} \right| = 1 + \frac{4mC_1 R^2 G_0 s + 2mC_2 r \cdot s + 2m^2 L C_2 s^2}{4 \left(1 - \frac{m^2-1}{4m} C_2 r s \right) + (1-m^2)L_2 C_2 s^2} \quad \dots\dots(31)$$

Putting $C_1 R^2 = L$, so that $C_1 R = L/R$, and then letting *R* go large and *r* small, we find for all three cases that the above formulae reduce to

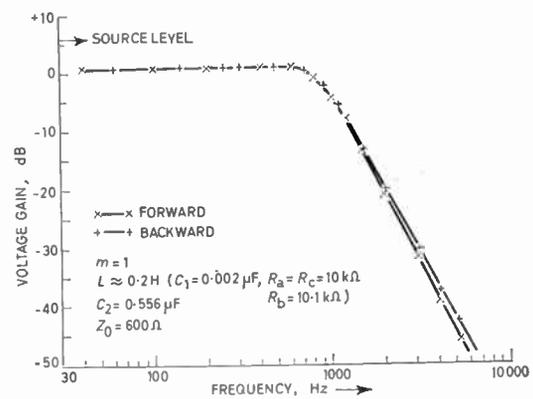
$$\left. \frac{v_1}{v_2} \right|_f \quad \text{and} \quad \left. \frac{v_1}{v_2} \right|_b \Rightarrow 1 + \frac{4mL G_0 s + 2m^2 L C_2 s^2}{4 + (1-m^2)L C_2 s^2} \quad \dots\dots(32)$$

The measured results for a 600 Ω filter with a cut-off frequency of 955 Hz are shown in Fig. 6 for $m = 1$ (prototype) at (a) and for $m = 0.6$ in Fig. 7. $C_1 = 0.002 \mu\text{F}$, $R = 10 \text{ k}\Omega$ and $L = 0.2 \text{ H}$, as before, and $C_2 = 0.556 \mu\text{F}$. The actual inductors used for comparison purposes were a Muirhead A-517-D for $m = 1$, together with a A-517-C for $m = 0.6$, both of which have a very low *Q* (about 5) at the cut-off frequency. Using 1% components the transmission characteristics were found to be very nearly the same in the two directions. The only part of the characteristic at which a significant difference occurred was in the vicinity of the attenuation peak in the *m*-derived case. At the peak there was a difference of about 10 dB. By very slight adjustment of R_b or R_c this difference could be removed and a common value of 29 dB was achieved; this occurred with R_b at 9.92 kΩ. Adjustment of R_b for maximum attenuation at the peak produced 45 dB attenuation. The passive coil had a fairly uniform extra loss of about 3 dB over most of the pass band, only 15 dB at the attenuation peak. However, at frequencies well beyond that of the attenuation peak all three characteristics were very nearly the same.

As an example of the dependence on component variation we show, for the prototype filter section, the forward and backward voltage transfer characteristic with $(R_b + 100) \Omega$ (i.e. 1% above its nominal value), which is on the side tending towards instability. It will be seen from Fig. 6(b) that there is a gain of nearly 1 dB in the pass-band and a reduction of approximately 250 Hz in the cut-off frequency, but the



(a) For $R_a = R_b = R_o = R$ in the inductor circuit (Fig. 3(a)).



(b) For R_b 1% above R .

The corresponding characteristic for an actual (passive) coil is included for comparison.

Fig. 6. Forward and backward voltage transfer-ratio characteristics of the prototype section operating between 600 ohm terminations.

basic curve shape remains the same. In the stop-band there is now a slight difference between the slopes of the cut-off characteristics for forward and backward transmission which causes the two curves to deviate increasingly with increase of frequency but this is not important.

5. Acknowledgment

The authors wish to acknowledge the U.K. Science Research Council's support for the work described in this paper.

6. References

- Holt, A. G. J. and Taylor, J., 'Method of replacing ungrounded inductors by grounded gyrators', *Electronics Letters*, 1, p. 105, June 1965.
- Sheahan, D. F., 'Gyrator flotation circuit', *ibid.*, 3, p. 39, January 1967.
- Holmes, W. H., Greutzmann, S. and Heinlein, W. E., 'Direct-coupled gyrators with floating ports', *ibid.*, 3, p. 46, February 1967.
- Ford, R. L. and Girling, F. E. J., 'Active filters and oscillators using simulated inductances', *ibid.*, 2, p. 52, February 1966.
- Riordan, R. H. S., 'Simulated inductors using differential amplifiers', *ibid.*, 3, p. 50, February 1967.
- Keen, A. W. and Peters, Jacqueline, 'Nonreciprocal representation of the floating inductor with grounded amplifier realisations', *ibid.*, 3, p. 369, August 1967.
- Schulz, H., 'Die Transformation der Vierpole-Kettenmatrix in die Diagonalform, eine Systematik aller linearen Vierpole und ihre Schaltungssymbolik', *Archiv der Elektrischen Übertragung*, 5, pp. 257-66, June 1951.

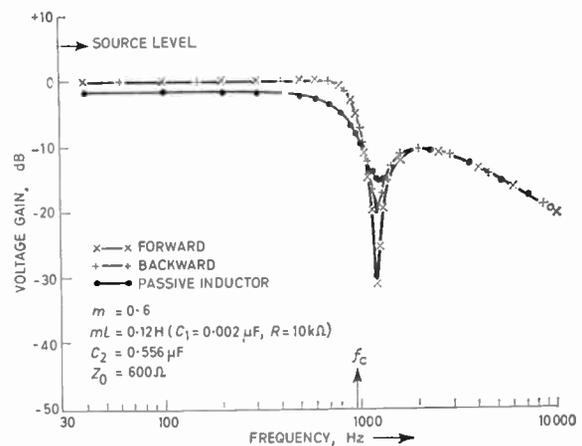


Fig. 7. As Fig. 6(a) for the shunt m -derived case.

- Keen, A. W., 'Immittance synthesis in the non-reciprocal domain', *Proc. Instn Elect. Engrs*, 110, p. 2118, December 1963. (I.E.E. Paper No. 4328E.)
- Llewellyn, F. B., 'Some fundamental properties of transmission systems', *Proc. Inst. Radio Engrs*, 40, p. 271, March 1952.
- Gewertz, C. M., 'Synthesis of a finite four-terminal network from its prescribed driving-point and transfer functions', *J. Math. Phys.*, 12, p. 1, January 1933.
- Fjällbrant, T., 'Activity and stability of linear networks', *Trans. Inst. Elect. Electronics Engrs on Circuit Theory*, CT-12, p. 12, March 1965.

Manuscript first received by the Institution on 4th January 1968 and in final form on 19th March 1968. (Paper No. 1199/CC14.)

© The Institution of Electronic and Radio Engineers, 1968

Of Current Interest . . .

New British Government Agency for Communications

Projects ranging from a simple preliminary survey to the provision of a complete Earth Station for satellite communications are to be undertaken by the new advisory and consultative agency called Crown Communications, set up by the Crown Agents in close association with the British Post Office. The services of the agency will be available to all countries throughout the world.

The new agency offers a complete range of services, from traffic and feasibility studies to the preparation of specifications for Earth Stations, tender recommendations, initial cost studies, financial advice and, if required, financial assistance on favourable terms. Project supervision, acceptance testing, commissioning, and the training of personnel for operation and maintenance duties are also available.

Crown Communications which has its headquarters at 4 Millbank, London, S.W.1, is independent of all manufacturing organizations and will be completely impartial in its advice and recommendations to overseas administrators.

Liquid Crystals

Liquid crystals—a strange, in-between kind of matter that until recently was largely a laboratory curiosity—are finding applications in industry. So named because, at one and the same time, they behave both as a liquid and as a solid, in appearance they are rather thick, colourless liquids. However, they act like solid crystals, for example, in scattering visible light. This scattering produces vivid colour changes whenever conditions shift the normal arrangement of the liquid crystals' molecules. The colours arise when white light shines on the crystals and the various colours making up the light are reflected at different angles by the cork-screw-like molecular structure of the material. Changes in temperature affect the twist of the spiral structure and change the colour reaching the eye.

Thus, with rise in temperature, the colour goes from red to yellow to green to blue and, finally, to violet. The colour change is not permanent, but shifts back again as the temperature falls. The response can be very sensitive. Less than one-tenth of a degree can cause the liquid crystals to run through their gamut of colours.

One of the first substances to be recognized as a liquid crystal was an organic chemical compound known as cholesteryl benzoate. When pure cholesteryl benzoate is heated it first melts to a viscous, cloudy liquid at 145°C with all the characteristics of a true melting point except that the liquid is not clear. The cloudiness remains until the temperature reaches 179°C; then the liquid suddenly clears. Thus cholesteryl benzoate appears to be a true (crystalline) solid only at temperatures below 145°C and a true liquid only at temperatures above 179°C.

Many substances in addition to cholesteryl benzoate have been found to behave as liquid crystals. Each

compound possessing such properties is characterized by a different pair or set of transition temperatures, just as ordinary compounds are characterized by different melting points. Mixtures of compounds as well as pure compounds have been studied by Westinghouse scientists in recent years.

Although known for many years, liquid crystals were 'rediscovered' at Westinghouse Research Laboratories in 1964. The Westinghouse insulating materials division now has developed the temperature-sensitive materials commercially for industrial use under the name of Spectratherm liquid crystal systems.

Typically, they are used for the non-destructive testing of metallic and non-metallic structures and for checking the operation of such devices as integrated circuits and transistors. The procedure is that the materials are simply sprayed or painted on a surface, where their colours show the temperature profile and pinpoint the location of 'hot spots' in electronic devices, cracks in metal structures, poor bonding between materials and similar faults.

Spectratherm liquid crystals have a response range, from red to violet, of from 1 deg C to 50 deg C. The ranges begin at essentially any temperature from about -25°C to 250°C. In use, the liquid crystals usually are painted on a black water-soluble undercoat to make the colour changes more visible.

In addition to changes in temperature, liquid crystals respond to other environmental changes such as pressure, chemical vapours and electric fields.

Other laboratories in the U.S.A., for example R.C.A., are investigating interesting applications in opto-electronics which liquid crystals could make feasible: these include displays, perhaps even as television picture tubes.

Agreement on Scientific Co-operation between Royal Society and the Czechoslovak Academy of Sciences

The Royal Society and the Czechoslovak Academy of Sciences concluded an agreement on scientific co-operation which came into force on 1st April last. The signatories were Professor P. M. S. Blackett, O.M., C.H., President of the Royal Society, and Professor F. Sorm, President of the Czechoslovak Academy of Sciences.

The agreement calls for both sides to facilitate contacts between specialists in the scientific disciplines within their mutual competence, and makes specific provision for visits of two weeks' duration by two senior scientists annually from each side, with the object of acquainting themselves with developments in their particular sphere in the host country, visiting scientific institutions, lecturing and holding seminars as may be appropriate, and generally informing their hosts of developments in their own countries.

In addition, provision is made for the annual exchange of up to three scientific research workers from each side to work in the other country on fundamental problems in physical and biological sciences.

Design of Resistively Mismatched Single Amplifier Stages

By

Professor

S. VENKATESWARAN, B.Sc.,
M.A., Ph.D., D.I.C., C.Eng., M.I.E.E.†

AND

V. P. NAMJOSHI, B.Tech.†

Summary: Resistively mismatched, reactively tuned single amplifier stages can be designed for a given measure of stability or performance factor, n , by the analytical solution of appropriate equations. The present paper provides a computer solution in the form of graphs for values of performance factor ranging between 1 and 10 and a closely accurate analytical solution for $n \geq 10$. Using these results, designs of mismatched single stages can be worked out. A design is worked out by way of example.

1. Introduction

An amplifier is an active network (a network that contains one or more 'electron device') that yields a maximum available power gain greater than unity. Both vacuum and semiconductor triodes are electron devices that have appreciable reverse feedback at higher frequencies, where reactive elements of device equivalent circuits cannot be neglected. A straightforward maximization of power gain may and usually does result in instability due to the nature of 'internal loop gain',

$$\frac{p_{12} p_{21}}{p_1 p_2} \text{ or } g_1$$

Here

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

is the matrix of the device network in h -, y -, z - or g -matrix environment.

$$p_1 = p_{11} + p_s, \quad p_2 = p_{22} + p_L$$

and p_s, p_L are the source and load immittances.

To make the amplifier stage stable for all values of the imaginary parts of ports' terminations, as happens during the process of tuning, $|g_1|$ must be reduced somehow to a value less than 1. There are two basic methods for accomplishing this. These are (i) unilateralization and (ii) mismatch.

In the first method, $|p_{12}|$ is reduced by the application of external passive feedback. If p_{12} of the 'composite' or 'resultant' network, i.e. of active plus passive feedback networks equals zero, the network is said to be 'unilateralized' while the process is known as 'unilateralization'.^{1, 2} When only the imaginary part of p_{12} is reduced to zero, the process is called 'neutralization'.

† Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India.

The real parts of p_1 and p_2 are artificially increased in the second method known as 'resistive mismatch' or just 'mismatch'.³ This prevents 'conjugate match' with its associated maximum available power gain and hence the name 'mismatch'. Increasing the real parts of p_1 and p_2 decreases

$$\left| \frac{p_{12} p_{21}}{p_1 p_2} \right| \text{ or } |g_1|$$

The maximum value of $|g_1|$ occurs when the imaginary parts of p_1 and p_2 both equal zero. The maximum loop gain is given by

$$|g_1|_{\max} = \frac{|p_{12} p_{21}|}{\text{Re}(p_1) \text{Re}(p_2)} = \frac{L}{\rho_1 \rho_2} = \frac{1}{n} \dots\dots(1)$$

where

$$p_{12} p_{21} = M + jN = L/\theta \dots\dots(2)$$

$$p_1 = \rho_1 + j\sigma_1; \quad p_2 = \rho_2 + j\sigma_2 \dots\dots(3)$$

and n is the performance factor.^{2, 4-6}

A particular performance factor is obtained when the real parts of source and load immittances vanish. Under these conditions

$$n_i = \frac{\text{Re}(p_{11}) \text{Re}(p_{22})}{|p_{12} p_{21}|} = \frac{\rho_{11} \rho_{22}}{L} = \frac{1}{|g_{1i}|_{\max}} \dots\dots(4)$$

n_i is called the 'inherent performance factor',^{2, 4-6} and g_{1i} the 'inherent loop gain' of the two-port network.

For a given measure of stability or performance factor, n , the operating power gain of the two-port network can be maximized. To find the associated terminations and maximum power gain, a cubic equation must be solved. This cubic equation has only one real root⁷ when $n \geq 1$. Once this real root is known (for $n \geq 1$) the maximum power gain and optimum terminations can be obtained. The curves given below have been obtained from solutions of

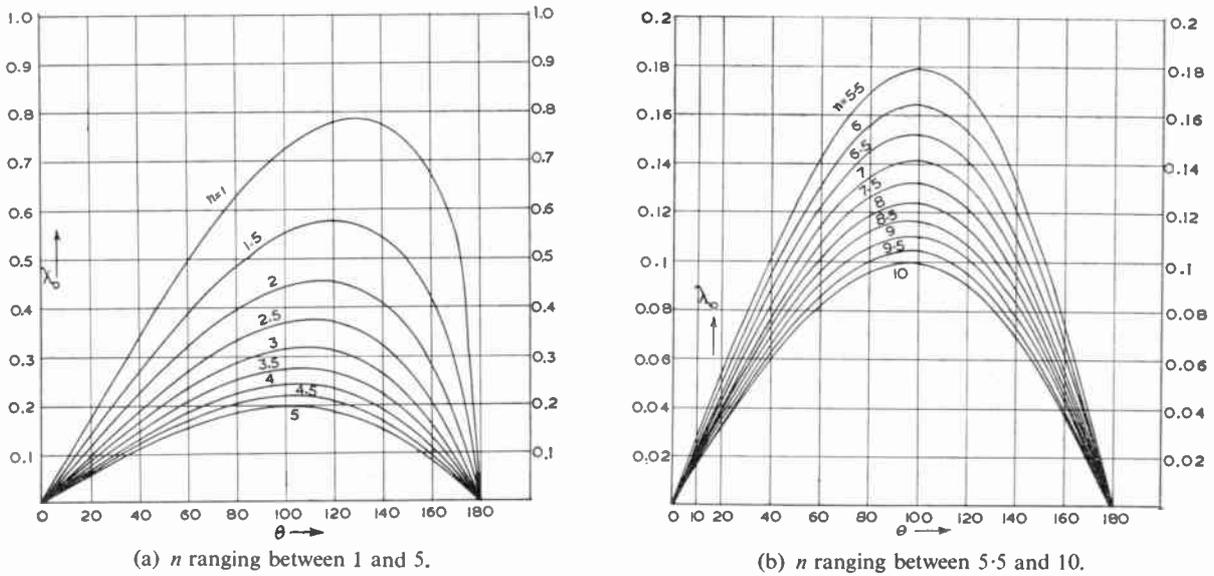


Fig. 1. Variation of λ_0 with θ for different values of performance factor, n .

appropriate equations using a computer. They cover a range of performance factor values between 1 and 10 and suggest an analytical solution when $n \geq 10$. Mismatched amplifier stages with $n \geq 1$ can now be designed straightaway.

2. Computer Solution

The operating power gain of a two-port network is given by

$$g = \frac{4|p_{21}|^2 \rho_S \rho_L}{D} \dots\dots(5)$$

where

$$D = |p_1 p_2 - p_{12} p_{21}|^2 \dots\dots(6)$$

must be minimized to give the maximum power gain.

From eqns. (2), (3) and (6)

$$D = |(\rho_1 + j\sigma_1)(\rho_2 + j\sigma_2) - M - jN|^2 = \left| \rho_1 \rho_2 \left(1 + j \frac{\sigma_1}{\rho_1} \right) \left(1 + j \frac{\sigma_2}{\rho_2} \right) - M - jN \right|^2 \dots\dots(7)$$

D is reciprocal and hence its maxima and minima with respect to σ_1, σ_2 will occur when

$$\frac{\sigma_1}{\rho_1} = \frac{\sigma_2}{\rho_2} = \lambda \text{ (say)} \dots\dots(8)$$

From eqns. (1), (2), (7) and (8)

$$\frac{D}{L^2} = \{n(1 - \lambda^2) - \cos \theta\}^2 + \{2n\lambda - \sin \theta\}^2 \dots\dots(9)$$

$\frac{d}{d\lambda} \left\{ \frac{D}{L^2} \right\}$ equated to zero gives the cubic equation

$$\lambda^3 + \left\{ 1 + \frac{\cos \theta}{n} \right\} \lambda - \frac{\sin \theta}{n} = 0 \dots\dots(10)$$

The only real root⁷ of this equation for $n \geq 1$, can be called λ_0 . Equation (10) has been solved for values of n from 1 to 10 in steps of 0.5, for angle θ ranging between 0° to 180° in steps of 10° . For angle θ between 180° to 360° or -180° to -0° , the root is obtained from

$$\lambda_0(-\theta) = -\lambda_0(\theta) \dots\dots(11)$$

by inspection of cubic eqn. (10) above.

The minimum value for D, D_{\min} , is obtained by multiplying eqn. (10) by λ and substituting in eqn. (9) with λ_0 replacing λ . Thus

$$\frac{D_{\min}}{L^2} = n^2(1 + \lambda_0^2) + n\{(\lambda_0^2 - 2) \cos \theta - 3\lambda_0 \sin \theta\} + 1 \dots\dots(12)$$

For a given value of the performance factor, n , the numerator of eqn. (5) can be maximized²; it is a maximum when

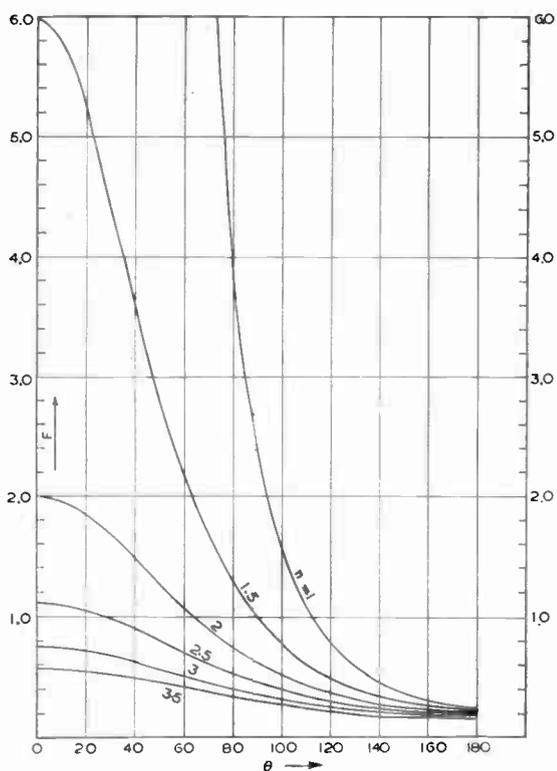
$$\frac{\rho_1}{\rho_{11}} = \frac{\rho_2}{\rho_{22}} = \sqrt{\frac{n}{n_i}}$$

or

$$\frac{\rho_S}{\rho_{11}} = \frac{\rho_L}{\rho_{22}} = \sqrt{\frac{n}{n_i}} - 1 \dots\dots(13)$$

while the denominator D is a minimum when

$$\sigma_S = \lambda_0 \rho_1 - \sigma_{11}; \quad \sigma_L = \lambda_0 \rho_2 - \sigma_{22} \dots\dots(14)$$



(a) n ranging between 1 and 3.5.

Fig. 2. Variation of F with θ for different values of performance factor, n .

Thus the maximum operating power gain for a given performance factor, n , is given by eqns. (5) and (13) as

$$g_{\max n} = \frac{4|p_{21}|^2 \rho_1 \rho_2 (1 - \sqrt{n_i/n})^2}{D_{\min}}$$

$$= F \times 4 \left| \frac{p_{21}}{p_{12}} \right| \left\{ 1 - \sqrt{\frac{n_i}{n}} \right\}^2 \dots\dots(15)$$

where

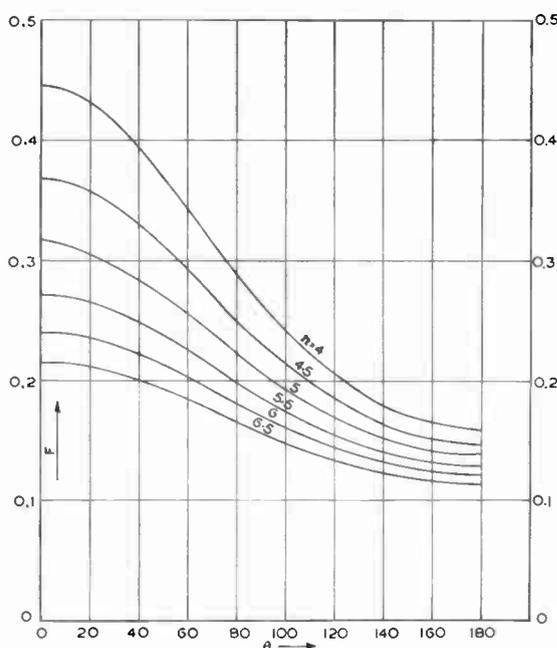
$$F = \frac{\rho_1 \rho_2 L}{D_{\min}} = \frac{nL^2}{D_{\min}} \dots\dots(16)$$

For given n ($n \geq 1$) and θ , solution of eqn. (10) gives the only real root λ_0 . Substitution of λ_0 in eqn. (12) yields the normalizing factor F of eqn. (16). Plots of λ_0 and F versus θ for various values of n are given in Figs. 1(a), 1(b) and 2(a), 2(b), 2(c) respectively.

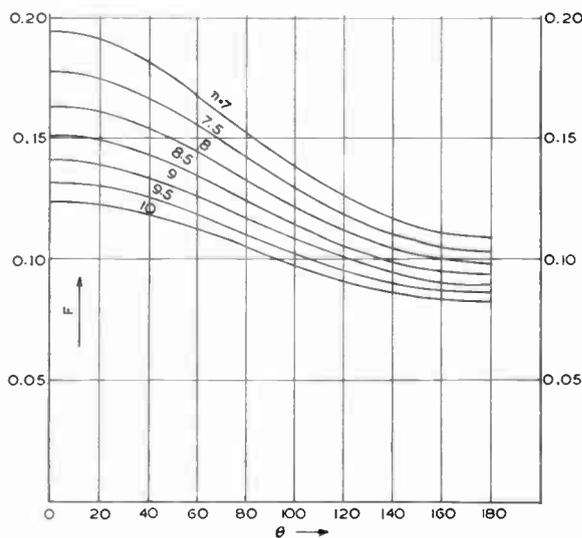
The terminations for obtaining the maximum power gain for a given n are given by eqns. (13) and (14).

3. Approximate Solution when $n \geq 10$

Plots of λ_0 and F versus θ in Figs. 1 and 2 cover the practical range of n values between 1 and 10. Yet, it



(b) n ranging between 4 and 6.5.



(c) n ranging between 7 and 10.

is of academic interest to know $g_{\max n}$ and optimum terminations for $n \geq 10$.

$g_{\max n}$ is given by the approximate expression^{2, 6} reported earlier, namely

$$g_{\max n} \approx \frac{4}{n} \left| \frac{p_{21}}{p_{12}} \right| \left\{ 1 - \sqrt{\frac{n_i}{n}} \right\}^2 \dots\dots(17)$$

The real parts of source and load terminations are given by eqn. (13). Equation (14) gives the imaginary

parts of these terminations provided λ_0 is known. For $n \geq 5$ the maximum value of $\lambda_0 \approx 1/n$ and this peak occurs at an angle θ which decreases from $\sim 100^\circ$ to $\sim 95^\circ$ as n increases from 5 to 10. The curve of λ_0 versus θ is approximately sinusoidal, with λ_0 value for $(\pi - \phi)$ being slightly greater than λ_0 value for ϕ , where $\phi < 90^\circ$. Thus Fig. 1(b) suggests that for $n \geq 10$,

$$\lambda_0 \approx \frac{\sin \theta}{n + \cos \theta} \quad \dots\dots(18)$$

Equation (10) can be rewritten as

$$\lambda^3 + \lambda \left\{ 1 + \frac{\cos \theta}{n} - \frac{\sin^2 \theta}{(n + \cos \theta)^2} \right\} - \frac{\sin \theta}{n} \approx 0 \quad \dots\dots(19)$$

where the coefficient of λ has been changed by a fraction

$$\frac{n \sin^2 \theta}{(n + \cos \theta)^3}$$

whose maximum value is $\sim 1/98$ for $n = 10$ and occurs for $\cos \theta \approx -0.15$. This is proved in the Appendix. For other angles or larger values of n , the error is even less.

Equation (19) can be factorized as

$$\left\{ \lambda - \frac{\sin \theta}{n + \cos \theta} \right\} \left\{ \lambda^2 + \left(\frac{\sin \theta}{n + \cos \theta} \right) \lambda + \frac{n + \cos \theta}{n} \right\} \approx 0 \quad \dots\dots(20)$$

It yields for the only real root λ_0 the value specified by eqn. (18).

Thus for $n \geq 10$, eqn. (14) becomes

$$\sigma_s \approx \left(\frac{\sin \theta}{n + \cos \theta} \right) \rho_1 - \sigma_{11} \quad \dots\dots(21)$$

and

$$\sigma_L \approx \left(\frac{\sin \theta}{n + \cos \theta} \right) \rho_2 - \sigma_{22} \quad \dots\dots(22)$$

The validity of these expressions in eqns. (17), (21) and (22) was further tested in the following example.

4. Illustrative Example

Consider a transistor type AF117 operating at 455 kHz in the common-emitter configuration for an emitter current of 2 mA, collector-base voltage of -6 V and an ambient temperature of 298°K. With biasing, a typical set of measured admittance parameters are

$$y_{11} = (510 + j 410) 10^{-6} \Omega^{-1}$$

$$y_{12} = -j 8 \times 10^{-6} \Omega^{-1}$$

$$y_{21} = 70 \times 10^{-3} \Omega^{-1}$$

and

$$y_{22} = (2.5 + j 14) 10^{-6} \Omega^{-1}$$

Therefore

$$y_{12} y_{21} = 0.560 \times 10^{-6} \angle -90^\circ \Omega^{-2}$$

The inherent performance factor, n_i , for this y -environment is given by eqn. (4) and the above admittance parameters as 2.28×10^{-3} . Let us design a mismatched amplifier stage for $n = 10$. For $\theta = -90^\circ$, the curves of Fig. 1(b) give $\lambda_0 = -0.099$ because of eqn. (11). Figure 2(c) gives the normalizing factor F as 0.102.

$$g_{\max n} = F \times 4 \left| \frac{p_{21}}{p_{12}} \right| \left\{ 1 - \sqrt{\frac{n_i}{n}} \right\}^2 = 35.4 \text{ dB}$$

From eqns. (13) and (14), the source and load terminations are obtained.

$$\text{or } \rho_s = 33.3 \text{ m}\Omega^{-1} \quad \sigma_s = -3.76 \text{ m}\Omega^{-1}$$

$$L_s = 93.1 \mu\text{H}\ddagger \quad \text{or } 92.3 \mu\text{H}\S$$

$$\rho_L = 163 \mu\Omega^{-1}; \quad \sigma_L = -30.4 \mu\Omega^{-1}$$

or

$$L_L = 11.5 \text{ mH}\ddagger \quad \text{or } 5.93 \text{ mH}\S$$

The mismatched stage has been designed with the reactances tuned to 455 kHz. These terminations are shown in Fig. 3, assuming a value of 10 pF for the total (coil + stray) capacitance at each port.

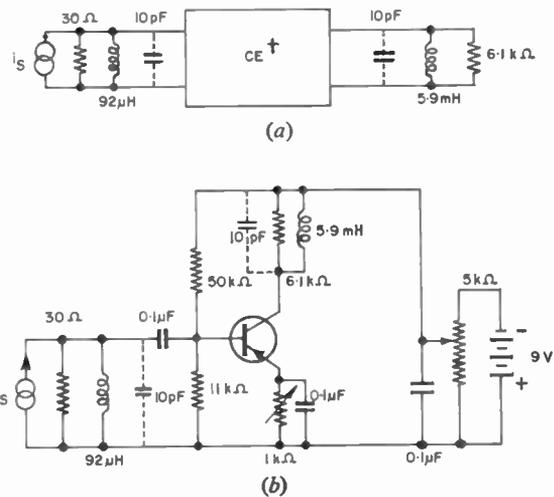


Fig. 3. (a) Terminations for a performance factor of 10 in the common-emitter configuration and y -matrix environment. (b) Complete circuit diagram of the mismatched amplifier stage. Total capacitance (coil + stray) is shown by dotted lines. Device used is a transistor type AF117 at $I_o = 2 \text{ mA}$, $V_{ob} = -6 \text{ V}$ at 298°K.

† Measured admittance parameters of 'black box' include the biasing resistances.

‡ Without any coil or stray capacitance.

§ With a realistic value of 10 pF for the total (coil + stray) capacitance.

Table 1

Exact and approximate maximum power gains for a typical AF117 in common-emitter configuration and y matrix environment

n	g_1 dB	g_2 dB	$(g_1 - g_2)$ dB
1	48.80	45.02	3.78
1.5	45.05	43.35	1.70
2.0	43.14	42.13	1.01
2.5	41.80	41.19	0.61
3.0	40.88	40.43	0.45
3.5	40.13	39.78	0.35
4.0	39.47	39.21	0.26
5.0	38.43	38.26	0.17
6.0	37.60	37.49	0.11
7.0	36.92	36.83	0.09
8.0	36.33	36.26	0.07
9.0	35.81	35.76	0.05
10.0	35.35	35.31	0.04

$$g_1 = F \times 4 \left| \frac{p_{21}}{p_{12}} \right| \left\{ 1 - \sqrt{\frac{n_1}{n}} \right\}^2$$

$$g_2 = \frac{4}{n} \left| \frac{p_{21}}{p_{12}} \right| \left\{ 1 - \sqrt{\frac{n_1}{n}} \right\}^2$$

Maximum power gains for n between 1 and 10 were calculated and are shown in Table 1 along with the gains obtained through the approximate expression of eqn. (17). The difference between exact and approximate gains are tabulated in the last column of the same table. It can be seen that for $n = 10$, the difference or error in gain is less than 0.05 dB. Therefore eqn. (17) is closely accurate for $n \geq 10$.

Values of λ_0 were calculated for $n = 10$ when θ varied from 10° to 170° in steps of 10° . These values are shown in Table 2 along with the values obtained by the use of eqn. (18). By either method, $\lambda_0 = 0$ for $\theta = 0^\circ$ or 180° . The agreement is very close.†

5. Acknowledgments

The authors are grateful to Dr. P. K. Kelkar, Director and Dr. H. K. Kesavan, Head of the Computer Centre, Indian Institute of Technology, Kanpur for the facilities provided. They are thankful to Mr. P. S. Sarma of the Electrical Engineering Department, I.I.T., Kanpur for his comments on this paper.

† Equation (18) is found to give values for λ_0 close enough to results obtained with the computer. The maximum error does not exceed 5% for value of n as low as 5. The maximum error does not exceed 1% for $n = 10$.

Table 2

Exact and approximate real roots of cubic equation for $n = 10$

θ in degrees	λ_1	λ_2	$\frac{\lambda_2 - \lambda_1}{\lambda_1} \times 100$
10	0.01580	0.01581	0.0506
20	0.03124	0.03126	0.0735
30	0.04593	0.04602	0.1848
40	0.05951	0.05971	0.3275
50	0.07163	0.07198	0.4880
60	0.08195	0.08248	0.6460
70	0.09015	0.09086	0.7890
80	0.09593	0.09680	0.9030
90	0.09903	0.10000	0.9780
100	0.09923	0.10022	0.9977
110	0.09637	0.09730	0.9610
120	0.09038	0.09116	0.7810
130	0.08129	0.08187	0.7130
140	0.06925	0.06961	0.5210
150	0.05456	0.05474	0.3300
160	0.03769	0.03775	0.1570
170	0.01925	0.01926	0.0572

$\lambda_1 = \lambda_0$ from computer results
[plotted in Fig. 1(b)]

$$\lambda_2 = \frac{\sin \theta}{n + \cos \theta}$$

6. References

- Cheng, C. C., 'Neutralization and unilateralization', *Trans. Inst. Radio Engrs on Circuit Theory*, CT-2, pp. 138-45, June 1955.
- Venkateswaran, S. and Boothroyd, A. R., 'Power gain and bandwidth of tuned transistor amplifier stages', *Proc. Inst. Elect. Engrs*, 106B, Suppl. 15, pp. 518-29, January 1960.
- Stern, A. P., 'Stability and power gain of tuned transistor amplifiers', *Proc. I.R.E.*, 45, pp. 335-43, March 1957.
- Venkateswaran, S., 'An invariant stability factor and its physical significance', *Proc. I.E.E.*, 109C, pp. 98-102, 1962. (I.E.E. Monograph No. 468E, September 1961.)
- Spence, R., 'On the latitude of choice of tuned amplifier terminations', *Trans. I.R.E.*, CT-9, pp. 336-9, December 1962.
- Venkateswaran, S., 'Performance factor of linear two-port active networks', *The Radio and Electronic Engineer*, 35, No. 2, pp. 109-14, February 1968.
- Venkateswaran, S. and Spence, R., 'Maximum power gain of an active two-port network', *Proc. I.E.E.*, 51, pp. 402-3, February 1963 (Letters).

7. Appendix: Maximum Error in the Coefficient of λ

The error, ϵ , in the coefficient of λ is given by

$$\epsilon = \frac{n \sin^2 \theta}{(n + \cos \theta)^3} \dots\dots(23)$$

It is a maximum or minimum with respect to θ , when

$$\frac{d\varepsilon}{d\theta} = \frac{2n(n + \cos \theta)^3 \sin \theta \cos \theta + 3n(\sin^3 \theta)(n + \cos \theta)^2}{(n + \cos \theta)^6}$$

$$= 0 \quad \dots\dots(24)$$

Since $n \geq 10$, $(n + \cos \theta)$ cannot be zero. Thus, for maxima or minima

i.e.
$$\left. \begin{aligned} \sin \theta &= 0 \\ \theta &= 0^\circ \text{ or } 180^\circ \end{aligned} \right\} \quad \dots\dots(25)$$

or
$$\cos^2 \theta - 2n \cos \theta - 3 = 0$$

i.e.
$$\cos \theta = n - \sqrt{n^2 + 3} \quad \dots\dots(26)$$

Equation (26) can be simplified as

$$\cos \theta \simeq n - n \left\{ 1 + \frac{3}{n^2} \right\}^{\frac{1}{2}}$$

$$\simeq -\frac{1.5}{n}, \text{ since } n \geq 10 \quad \dots\dots(27)$$

$\theta = 0^\circ$ or 180° gives a minimum value for ε of 0

$$\cos \theta \simeq -\frac{1.5}{n}$$

gives the maximum value,

$$\varepsilon_{\max} \simeq \frac{n \left(1 - \frac{2.25}{n^2} \right)}{\left(n - \frac{1.5}{n} \right)^3} \quad \dots\dots(28)$$

For $n = 10$, $\varepsilon_{\max} \simeq 1/98$.

*Manuscript first received by the Institution on 17th July 1967 and in final form on 13th February 1968.
 (Paper No. 1200/CC15.)*

© The Institution of Electronic and Radio Engineers, 1968

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10} , from nominal frequency for June 1968

June 1968	24-hour mean centred on 0300 U.T.			June 1968	24-hour mean centred on 0300 U.T.		
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz
1	- 299.9	+ 0.1	0	17	- 299.9	—	0
2	- 299.9	+ 0.1	0	18	- 299.8	—	0
3	- 299.9	+ 0.1	0	19	- 300.0	—	0
4	- 299.9	0	0	20	- 299.9	—	0
5	- 299.9	+ 0.1	- 0.1	21	- 299.9	—	0
6	- 299.9	+ 0.1	0	22	- 299.9	+ 0.1	- 0.1
7	- 299.9	+ 0.1	- 0.1	23	- 300.0	0	- 0.1
8	- 300.0	+ 0.2	0	24	- 300.0	—	+ 0.2
9	- 300.0	0	- 0.1	25	- 300.0	0	+ 0.3
10	- 300.0	+ 0.1	0	26	- 300.0	+ 0.1	0
11	- 299.8	+ 0.1	0	27	- 299.9	+ 0.1	0
12	—	—	0	28	- 299.8	+ 0.1	0
13	—	+ 0.3	0	29	- 300.0	0	0
14	- 299.9	+ 0.2	0	30	- 299.9	+ 0.1	0
15	- 299.9	+ 0.1	+ 0.1				
16	- 299.8	—	0				

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.

Notes: (1) All measurements are made in terms of H.P. Caesium Standard No. 134 which agrees with the NPL Caesium Standard to 1 part in 10^{11} .

(2) The MSF 60 kHz transmission was suspended over the period 16th to 21st June due to work in progress at the transmitter.

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.

GUNN-EFFECT PULSE CODE MODULATION

Pulse code modulation represents a major contribution to the art of communication. Each modulating wave is sampled periodically at a rate somewhat in excess of twice its highest frequency component. The samples are quantized into discrete steps. The quanta are then expressed by a particular code pattern, which gives rise to a pattern of coded pulses. At the receiving end, each pulse pattern is decoded and made to produce a voltage proportional to the original quantized sample. A succession of such pulse patterns enables one to reconstruct the original wave. This modulation technique has two important properties. Firstly, it is possible to obtain a high quality signal under conditions of noise and interference so bad that it is just possible to recognize the presence of each pulse. Secondly, by using regenerative repeaters which detect the presence or absence of pulses and then emit reshaped pulses, the initial signal-to-noise ratio can be maintained through a long chain of repeaters.

In a paper written by a German engineer on work carried out at Sheffield, the application of Gunn-effect pulse devices is described for pulse code modulation, which can be employed for telecommunications and transformation of analogue into digital information. The presented modulator operates at least two orders of magnitude faster than modern devices in classical circuitry.

'Gunn-effect pulse code modulation', H. L. Hartnagel, *Archiv der Elektrischen Übertragung*, 22, No. 5, pp. 225-9, May 1968.

USE OF INTEGRATED CIRCUITS FOR ROCKET CONTROL SYSTEMS

Requirements exist in scientific sounding rockets for the control of a variety of operations which must be programmed to occur at specified times or in a particular sequence. These operations include extension or separation of experiment packages from the nose cone, power control of experiments, and time-sharing data transmitted by telemetry.

The number of programmed in-flight control operations increases with the number and complexity of experiments carried by a rocket. While control of small payloads which contain only one or two experiments can usually be done by simple mechanical timers or electronic delay circuits, more extensive program control is desirable in larger payloads than can conveniently be accomplished by a simple controller.

The availability of low-cost, packaged integrated circuits has made relatively complex control systems more attractive

in sounding rocket applications where size, power consumption, and reliability are important considerations. Size weight, and frequently power, can be reduced by using integrated circuits as compared with a discrete component design. Since the number of external electrical connections to be made is greatly reduced and since each circuit functional unit is batch fabricated and pretested by the manufacturer, a design with integrated circuits is also likely to result in a system with higher reliability.

In a Canadian paper the advantages of using integrated circuits in control applications in sounding rockets are discussed with reference to a timer designed for controlling the operation of a group of cameras.

'A role for integrated circuits in sounding rocket payload control applications', K. A. Steele, *Canadian Aeronautics and Space Journal*, 14, No. 2, pp. 57-60, February 1968.

THRESHOLD SENSITIVITY OF PHOTODIODES

The quality of a photodiode as a detector of weak light signals (or of small signal increments) is characterized by the amount of luminous power incident on the photodiode window that is just detectable. The threshold power W_0 is usually taken to be that amount of light flux, producing a response at the photodiode output, whose power is equal to the noise power at the photodiode output within the frequency bandwidth being measured (it is usually calculated for a 1 Hz bandwidth).

In the work described in a Russian paper the threshold powers of alloyed germanium and silicon photodiodes are determined for a light modulation frequency of 1 MHz. In order to determine the optimum operating conditions, the dependence of the photodiode threshold power on the magnitude of the negative bias applied to the photodiode and on the d.c. illumination level is studied.

On the basis of the measurements that have been made it is concluded that at frequencies of the order of 1 MHz the alloyed germanium and silicon photodiodes with a fairly thick base (about 100 μm) are far from ideal (noise figure of $F = 1$). In selecting the optimum mode of operation, the possibility of reducing the threshold power by the addition of d.c. illumination should be taken into consideration. One can expect that thin-base, diffused photodiodes with low dark current will have considerably smaller threshold power values at high light modulation frequencies.

'Measurement of threshold sensitivity of photodiodes at 1 MHz', I. A. Gavrilov and N. N. Kolachevskiy, *Radio Engineering and Electronic Physics* (English language edition of *Radiotekhnika i Elektronika*), No. 10, pp. 1683-6, October 1967,

TELEVISION COLOUR STANDARDS CONVERSION

The inclusion of American NTSC colour programme in the European television network operating with the PAL system necessitates (besides the conversion of the 525 lines/60Hz standard into 625 lines/50Hz, as is being done by black-and-white programmes) the transcoding of the colour information.

A German paper deals with the methods of transcoding which allows the two-parameter colour information to be converted into the new standard with a single converter. The methods allow electronic as well as optical converters to be used. The principle of the converter consists in converting the dotted carrier structure of a NTSC picture into a pattern consisting of vertical strips, which allows the necessary interpolation of the signal values of consecutive lines to be made during sampling at the output synchronizing standard.

'Methods of converting NTSC and PAL colour standards among television standards of different vertical frequencies', F. Jaeschke, *Nachrichtentechnische Zeitschrift*, 21, No. 4, pp. 177-81, April 1968.

IMAGE-CONVERTER TUBE APPLICATIONS

The use of image converter tubes for ultra-rapid photography enables exposure times to be reduced to below 10 ns. Moreover, they show an increase in brilliance of above 10, whereas other devices absorb a considerable proportion of the light. The shutter tube described in a French paper comprises a photo-cathode and a screen placed a few millimetres apart. It has enabled images of good quality to be obtained at exposure times of 1 ns. The image achieved a resolution of 20 pairs of lines per millimetre, showing no distortion.

'Shutter tube for ultra-rapid photography', C. Eschard and R. Polaert, *l'Onde Electrique*, 48, No. 494, pp. 426-9, May 1968.

OPTICALLY-COUPLED TRANSISTORS

Since the discovery of high-efficiency light emission from p-n junctions of GaAs, several new related devices have been proposed. Among these is the optically-coupled transistor, which combines a GaAs diode with a silicon photodiode. The optically-coupled elements may be used as electrically insulated control elements, semiconductor isolators, and optical coupling elements in solid-state circuits. In these applications, the emitted light may be modulated by a signal or the characteristics of its propagation may be controlled by a signal.

Optically-coupled transistors may be realized in many forms. The practicality of such a device depends on the time-response and the efficiency that can be attained without destroying its inherent advantage of electrical isolation.

As a fundamental study to evaluate its performance, the combination of a high-speed photodiode and a GaAs luminescent diode are investigated and the results are described in a Japanese paper. A general method of optimizing the response and efficiency of p-i-n photodiodes is derived. An experimental diode was tested for its pulse

response using a GaAs diode laser, and a time response of 10^{-10} s was obtained.

The absorption loss of a luminescent diode is evaluated considering the emission spectrum and the wavelength dependence of the absorption coefficient. An internal quantum efficiency of 17% is obtained. Techniques of reducing the reflection loss are studied. It is shown that optically-coupled transistors may be constructed to have an efficiency of a few per cent and a time response of 10^{-9} s using luminescent diodes.

'A fundamental investigation to evaluate optically coupled transistors', S. Orihara and H. Yanai, *Electronics and Communications in Japan* (English language edition of *Denki Tsushin Gakkai Zasshi*), 50, No. 3, pp. 94-102, March 1967.

RATE AND ACCELERATION MEASURING PENDULUM

This inertial navigation system, developed by a Swedish company primarily for aircraft navigation, is based on indication of changes of the vertical with a Schuler pendulum, whose oscillation time is 84.4 min. This device, which contains a gyro with integrating feedback, has been named RAMP (Rate and Acceleration Measuring Pendulum). The complete system comprises two RAMP units, mounted on an inertial platform, which is slaved to the pendulums. For the indication of heading, a directional gyro is mounted on the same platform. The equipment was tested both in the laboratory and during a number of test flights, including flights over the European continent and transoceanic flights. The average uncertainty amounted to about 2 to 3 nm per hour.

'The RAMP inertial navigation system', F. Hector, *Philips Technical Review*, 29, Nos. 3-4, pp. 69-85, 1968 (in English).

SEMICONDUCTOR DIODES AS MICROWAVE GENERATORS

Under controlled avalanche-breakdown conditions semiconductor diodes exhibit negative resistance suitable for microwave generation due to characteristic space-charge and transit-time effects.

Experimental and theoretical investigations of planar silicon diodes with abrupt and hyper-abrupt p-n junctions at high avalanche current densities are reported in a German paper. Quasi-stationary negative resistances and instabilities at about 100 MHz are observed which can be explained by a specific space-charge feedback for certain relations between avalanche density, doping density and length of the depletion layer. At microwave frequencies oscillations are produced at 1.3 to 4.3 GHz and 6.6 to 10.9 GHz. They are attributed to negative resistances measured under small-signal conditions. The upper interval is interpreted as the transit-time range for frequencies exceeding the avalanche resonance frequency. A cascade model is introduced to explain a space-charge delay mechanism which produces a negative resistance at very small transit angles for certain multiplication distributions and direct-current densities.

'Negative resistance and microwave generation with silicon diodes at high avalanche current densities', B. Höflinger, *Archiv der Elektrischen Übertragung*, 22, No. 2, pp. 87-101, February 1968.