THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 29 FEBRUARY 1965 NUMBER 2

SIR WINSTON CHURCHILL, K.G., P.C., O.M., C.H., F.R.S.

1874–1965

THE death of the greatest figure of this century must be marked by a brief note on the relationship of Sir Winston Churchill to the world of science and in particular to radio and electronic engineering. Writing in "The Second World War" he said "I know nothing about science but I know something about scientists, and had much practice as a Minister in handling things I did not understand". This recognition of how to handle the stream of new inventions and developments which was such a feature of "The Wizard War", as he termed it, was as vital a contribution as his control of the strategic conduct of military operations.

Churchill's association with the services during the 1914–18 War is well known, but many members may not be aware that, when Secretary of State for War and Air in 1920, he signed the Royal Warrant giving the Sovereign's approval to the formation of the 'Corps of Signals', now the Royal Signals.

Throughout the late 'thirties Churchill was kept in touch with the development of radar, serving on the Committee of Imperial Defence on Air Defence Research from 1935. His history of the war lays great stress on the vital role of radar, both in the Battle of Britain and later on in the period 1942–43 when blind bombing aids such as 'Gee','Oboe' and 'H_2S' were beginning to come into operation. Another important device, the radio countermeasure known as 'Window' (strips of metallized paper dropped from aircraft to act as tuned dipoles and confuse enemy radar), was in fact brought to the notice of the Air Defence Research Committee by Churchill in 1937 at the suggestion of Professor Lindemann. Even after final development in 1942 it called for direct intervention a year later by the Prime Minister in order to resolve the conflicting views of Fighter and Bomber Commands on its suppression or use.

Undoubtedly Churchill had genius in drawing out the best course of action from the proposals of committees of experts. Then came the further virtue of making sure that the decision was carried through irrespective of difficulties. Many members of this Institution will no doubt recall 'crash programmes' for producing first prototypes and then production models at the direct instructions of the Prime Minister.

Similar collaboration between government and science which Churchill fostered so ably in war could well lead in peacetime to the realization of the benefits of invention. In our field he had, as is recorded in the Institution's History, a conviction that the peaceful applications of radio could be a blessing and a defence.

F. W. S.

INSTITUTION NOTICES

Paris Component Show

The 1965 Salon International des Composants Electroniques is being held in Paris from 8th to 13th April. Concurrently with this will be an Electro-Acoustic Exhibition and an International Colloquium on "Techniques of Memory Circuits."

This Salon is the biggest exhibition of its kind in Europe and it is therefore appropriate that the Institution should take a stand in the section devoted to the technical press, at which information on *The Radio and Electronic Engineer* and other Institution publications will be available.

Members of the Institution residing in France and neighbouring countries are invited to call at the I.E.R.E. stand and it is hoped to arrange an informal meeting during the course of the Salon. Further details of such a meeting will be sent to members in France at a later date.

Conference on Thermionic Electrical Power Generation

An International Conference on Thermionic Electrical Power Generation is being organized under the auspices of the O.E.C.D. European Nuclear Energy Agency by the Institution of Electrical Engineers. The Conference, which is the first international meeting on this topic to be held in Europe, will take place at the I.E.E. headquarters in London during the week of 20th to 24th September, 1965.

Subjects to be dealt with are:

Convertor performance; theoretical considerations; surface phenomena; measurement techniques and diagnostics; in-pile studies; construction materials (e.g. nuclear fuels, emitter and collector materials, insulators and corrosion problems); applications (nuclear reactor, solar, isotope and flame heating); and new developments (heat pipes and ion injection).

Reports of work carried out in United States, U.K., French, German and Euratom research establishments will be presented, and further contributions are invited. Provisional titles and synopses of papers for the conference should be sent before 30th April to the Conference Secretariat, Institution of Electrical Engineers, Savoy Place, London, W.C.2.

Electronics in East Africa

Recently the Electronics Group of the East African Institution of Engineers staged a Radio and Electronics Show in Nairobi. It was the biggest exhibition of its type ever held in East Africa and by any standards achieved a very considerable success. At the official opening a message of goodwill from Mr. Jomo Kenyatta, the Prime Minister of Kenya, was sent out from the exhibition hall and as a result of a gigantic radio link-up was made to encircle the world.

Dr. J. G. Kiano, Minister of Commerce and Industry, who opened the show on behalf of the Prime Minister, pressed a button in the Gloucester Hall, Nairobi, which initiated the transmission of the message by land line to the East African External Telecommunications station at Ngong Hills, where it was retransmitted by radio to Singapore, and then by radio again to Sydney, Australia. From Sydney it was retransmitted by the new COMPAC submarine telephone cable to Auckland, New Zealand, and then across the Pacific ocean via Fiji and Honolulu to Vancouver. It was retransmitted across Canada by microwave relay, and then across the Atlantic Ocean by submarine cable to London, and finally across Europe and Northern Africa back to Nairobi by a mixture of radio and land line. The message was printed out on the terminal equipment in the hall itself after a time interval of 65 seconds.

The Chairman of the recently-formed Electronics Group of the East African Institution of Engineers is Dr. Denis Taylor (Member) who is in charge of the UNESCO Mission to the University College, Nairobi. Dr. Taylor served on Brit.I.R.E. Convention Committees in 1957 and 1961.

S.I.M.A. Exhibition in Moscow

For the fifth time the Scientific Instrument Manufacturers' Association are holding a group exhibition in Moscow. The agreement between S.I.M.A. and the All-Union Chamber of Commerce has been signed and the exhibition will be held in the Polytechnic Museum, Moscow, from 26th May to 4th June. It will feature the latest technical achievements of 32 S.I.M.A. members.

Astronautical Congress

The Sixteenth International Astronautical Congress will be held in Athens from 12th to 18th September, 1965.

There will be one main session on satellite laboratory systems; a session on general subjects connected with astrodynamics, meteorological and communication satellite systems, bioastronautics, life support systems, propulsion, and guidance and control, with particular reference to radiation problems; and special sessions on the Lunar International Laboratory, Astronautics in Education, and a Space Law Colloquium.

Offers of papers should be sent to the International Astronautical Federation, 250 Rue Saint-Jacques, Paris 5e.

U.D.C. 621.397.3

An Experimental Frame Difference Signal Generator for the Analysis of Television Signals

By

A. J. SEYLER†

This paper, originally presented at the 1963 Convention of the Institution of Radio Engineers Australia in Melbourne in May 1963, is reprinted from the November 1963 issue of the Proceedings of the Institution of Radio Engineers Australia. On the recommendation of the Institution of Electronic and Radio Engineers it has been awarded the 1964 Norman W. V. Hayes Memorial Medal of the Australian Institution.

Summary: One proposed method of reducing the channel capacity requirements for the transmission of television signals over long distances is known as frame difference signal (f.d.s.) coding. This method implies the recoding of television signals so that only those parts of every frame are transmitted which are different in two consecutive frames. In order to assess the possible savings in channel capacity resulting from this technique, it is first necessary to measure and record the probability density of f.d.s. areas in actual television programme signals.

In this paper an f.d.s. generator is described which, by using a vidicon camera tube as storage and subtracting device, produces frame difference signals suitable for of statistical analysis. The generated signals are further processed to obtain the frame by frame integral of the f.d.s. areas, and the ancillary equipment required to carry out this operation is described. The data are recorded for subsequent evaluation.

1. Introduction

Due to the fact that large portions of television images do not change in consecutive frames, it must intuitively be expected that a substantial correlation exists between the same spatial picture elements in consecutive frames.

Even before the formulation of Shannon's statistical theory of communication (information theory) the exploitation of this correlation had been suggested as a means of reducing the channel capacity required for the transmission of television signals[1,2] by transmitting only the difference between picture elements in consecutive frames, i.e. the frame difference signal (f.d.s.). By information-theory analysis it was then confirmed[3,4] that the application of predictive coding methods based on interframe correlation should yield substantial reductions of the required channel capacity.

In all cases, however, formidable difficulties seemed to be encountered in translating the theoretical findings into practice. This is due to the various causes from which f.d.s. can arise, usually from:

(i) movement in the picture,
(ii) movement of the picture (resulting from camera movement), and
(iii) scene-changes.

† Research Laboratories, Postmaster-General's Department, Melbourne, Victoria.

In order to apply statistical coding methods efficiently, the signals to be encoded are required to be statistically stationary stochastic processes. However, the physically different modes of f.d.s. origin introduce such large variations in the density of f.d.s. that the redistribution stores (queues) required[5] to average out these variations in time would become unmanageably large. In this context the term f.d.s.-density d_D is defined by the ratio of the number of Nyquist samples n_D which are different in consecutive frames to the total number n_F of Nyquist samples contained in one frame.

$$d_D = \frac{1}{n_F} \sum_F n_D \qquad \text{.....(1)}$$

where the Nyquist sampling interval t_N is given by the sampling theorem as

$$t_N = 1/2f_C \qquad \text{.....(2)}$$

f_C being the maximum frequency component in the waveform of the original television signal.

The difficulties arising from the large variations of the f.d.s.-density may be overcome by introducing coding methods which are adaptively matched to the state of the input signal. These operate on the basis that the average information content of the samples which are different in consecutive frame varies also because, if the f.d.s.-density becomes large due to movement of the picture (e.g., camera panning) the inherent storage properties of the optical-electrical

transducer (camera tube) cause 'blurriness' of the contours of moving objects. This movement-dependent low-pass filter effect then reduces the specific sample density (samples per unit area with respect to time) in the frame difference signal. Thus the average rate of information generated by the signal source will not increase proportionally with the f.d.s. density.

On the other hand scene changes may occur between consecutive frames in which *a priori* no movement is present and thus the source mechanism just described would not apply, the f.d.s. density would be quite large and so would be the specific sample density. However, considering the definition of 'information' as a property of the source-user pair[6] and applying the dynamic behaviour of the human sense of vision observable on scene changes[7,8,9] as a controlling coding criterion, it is found that the specific sample density may be substantially reduced in the first frame following a scene change, since the human receiver is not capable of recognizing fine detail in a newly presented scene for some few hundreds of milliseconds. Therefore again when on a Nyquist sampling basis alone the f.d.s. density would become large, the actual signal sample density will not increase proportionally.

The number of samples per frame of the f.d.s.-coded signal thus depends on the product of two inversely related quantities and therefore the fluctuations of the bit-rate should be substantially reduced.[10] In consequence it may be expected that a redistribution store of just sufficient capacity for averaging the intra-frame sample fluctuations will provide a satisfactory buffer between coder and channel and thus yield an efficient channel capacity utilization with economical means.

The previously envisaged difficulties in applying f.d.s. coding had exercised a restraining influence on the efforts devoted to measuring actual f.d.s. statistics. Therefore only a very small number of image frames were evaluated by photographic methods,[3] or the inter-frame correlation of one element only was measured in real time over a number of frames and for different programme material.[11] Neither of these suffice to make quantitatively valid estimates of the attainable channel capacity reductions.

An instrumentation was therefore devised by which f.d.s. densities and their fluctuations in real time could be measured on actual television programme material. It is the purpose of this paper to describe the method and equipment used for this investigation. Although some sections, particularly those required for the large-scale recording of f.d.s. data in computer language are still not completed, early results will be demonstrated here which may give an insight into the dynamics of f.d.s. statistics.

2. System Functions and Components

The functions and components of the overall system are best described by reference to Fig. 1, which shows that the system functions may be logically divided into:

(i) f.d.s. generation,

(ii) processing, and

(iii) recording.

A fourth section is indicated which will later deal with the processing of the actual video signal after being gated by the f.d.s. samples.

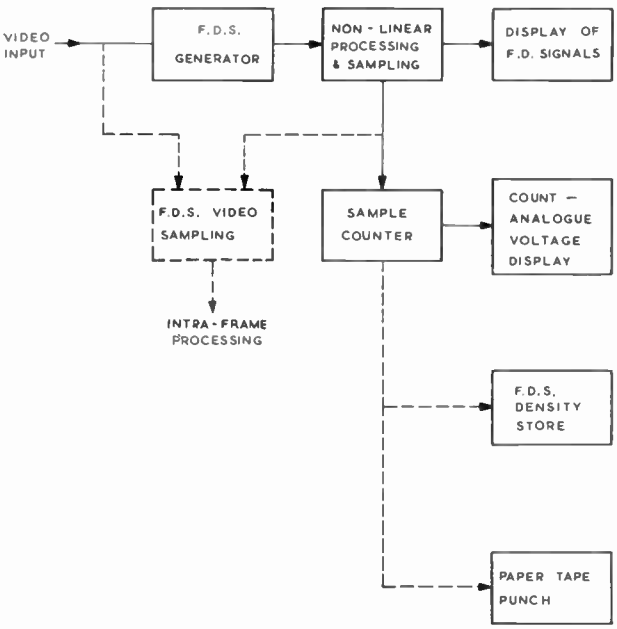


Fig. 1. Functional system schematic.

It is the purpose of the f.d.s. generator to generate the f.d.s. by subtracting two consecutive frames of the incoming video signal, i.e. carry out the operation described by eqn. (3).

$$f(t_k) - f(t_k - T) = g_D(t) \qquad \ldots\ldots(3)$$

where $f(t_k)$ represents the video signal waveform at time t_k, T equals one frame interval (40 msec), and $g_D(t)$ represents the f.d.s. waveform. The 'raw' f.d.s. thus obtained contains positive and negative going waveform sections, the amplitude and polarity of which will depend on the actual amplitude difference between the same spatial picture points in consecutive frames, the zero axis obviously obtaining for f.d.s. = 0. This signal is inevitably contaminated by random noise, partly already present in the input signal and partly originating in the f.d.s. generator. In addition, imperfections of the delay and subtracting device contained in the f.d.s. generator will cause spurious

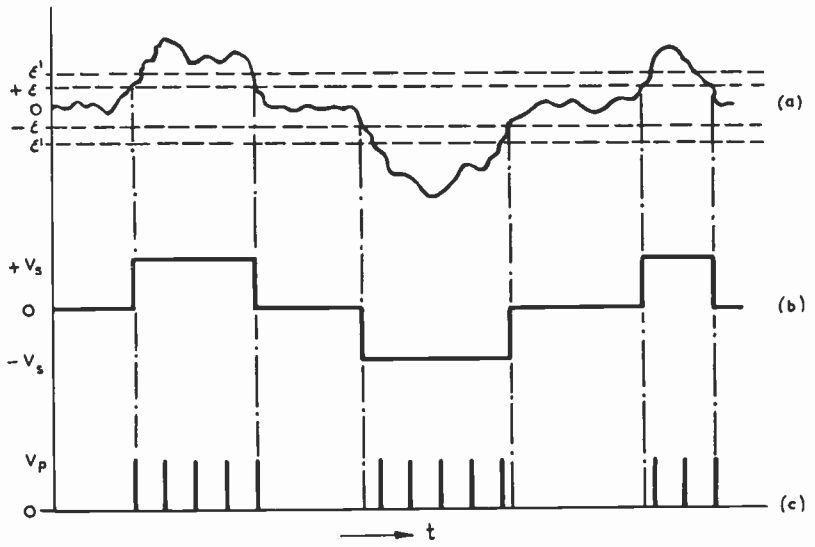


Fig. 2. Non-linear f.d.s. processing.
(a) Raw f.d.s.
(b) Sliced f.d.s.
(c) Nyquist sample pulses.

signals. Although for the purpose of statistical analysis of f.d.s. densities, noise can to some extent be accounted for in the evaluation process, it is nevertheless desirable to minimize its effects, a function which is incorporated in the non-linear processing and sampling section.

Since the immediate purpose of the system described here is the recording and analysis of f.d.s. densities (eqn. (1)) the raw f.d.s. must be transformed into constant-amplitude Nyquist samples n_D for which the condition applies.

$$n_D = 1 \quad |g_D(t)| \geqslant \varepsilon \qquad \text{.....(4a)}$$

$$n_D = 0 \quad |g_D(t)| < \varepsilon \qquad \text{.....(4b)}$$

In eqns. (4a) and (4b) ε is the decision threshold amplitude of the f.d.s. above which Nyquist sample-pulses are generated by the processing device. This threshold action is qualitatively demonstrated in Fig. 2. The raw f.d.s. (Fig. 2(a)) after being symmetrically clipped at the level $\pm \varepsilon'$, shaped and normalized to $\pm V_s$ (Fig. 2(b)) operates as gating signal which permits pulses to be obtained having a pulse repetition frequency of $t_N^{-1}\,s^{-1}$ and constant amplitude V_p (Fig. 2(c)). This gating action is at the same time subject to a threshold ε, so that no pulses can pass, if the initial f.d.s. amplitude is below this level, as indicated in Fig. 2(a). Thus in this process $g_D(t)$ is virtually sliced between ε and ε', and denoting this sliced and normalized signal by $\bar{g}_D(t,\varepsilon)$ the signal consisting of the sequence of sample pulses (Fig. 2(c)) may be expressed[12] by

$$n_D(t) = |\bar{g}_D(t,\varepsilon)|P_N(t) \qquad \text{.....(5)}$$

where

$$P_N(t) = \sum_{-\infty}^{+\infty} \delta(t-kt_N) \qquad \text{.....(6)}$$

is the series of narrow sampling pulses, periodically occurring at Nyquist intervals kt_N. No more rigorous theoretical treatment of this sampling mechanism is considered necessary here, since only the number of pulses, not their shape, is of importance in this context and the sampled signal is an on/off type signal only.

The pulse-signal so generated is then, for qualitative assessment of the f.d.s. generating and processing function, band-limited and displayed on the screen of a television monitor, resulting in 'f.d.s.-pictures' of which the screen-photographs shown in Fig. 3 are typical examples. To obtain quantitative values of f.d.s. densities and their fluctuations from frame to frame in time, $n_D(t)$ is fed to a sample counter. In this counter all samples generated during one frame interval T are counted and at the end of each frame interval the counter is reset to zero. Thus the states of the counter at

$$t_k = t - kT$$

represent, when normalized by multiplication with $1/n_F$, the time-variant f.d.s. density which in analogy to eqn. (1) becomes

$$d_D(t_k) = \frac{1}{n_F} \sum_{T} n_D(t) \qquad \text{.....(7)}$$

where $n_D(t)$ is defined by eqn. (5). (This sample counter can also be considered as an 'ideal integrator' having an integration time of one frame interval.)

Since the discrete values $d_D(t_k)$ of $d_D(t)$ arise at 40 ms intervals provision for permanently recording $d_D(t_k)$ must be made. For an initial survey and to obtain some preliminary results the digital output signal of the counter is converted into an analogue

(a) Lady in conversation.

(b) Two men.

(c) Horses with riders.

Fig. 3. Frame differences in television pictures.

voltage signal, displayed on a c.r.o. and photographically recorded. Typical records obtained this way are depicted in Fig. 4. (See Section 4.)

For an extensive statistical analysis the evaluation of such records would be considered too laborious. Therefore the digital signal representing $d_D(t_k)$ will eventually be recorded on paper tape by means of a high-speed tape-punch, for subsequent computer operations. As an intermediate step a multichannel digital accumulator is being assembled, containing a magnetic core memory in which up to $(2^{16}-1)$ occurrences of each of 2^m possible values of $d_D(t_k)$ may be stored, and thus a density histogram of $p(d_D, i)$ for $i = 1, 2, 3, \ldots 2^m$ may be obtained directly.

The f.d.s. gated sampling pulses will in a further extension of the project be used to gate the original video signal, so that only those video signal samples will be obtained which are different in a particular frame and are therefore necessary to be transmitted for replacement in the picture being re-synthesized in the receiving terminal. These samples would be subject to intra-frame processing methods to eliminate adjacent element redundancies, using, for example, a method as described by Cherry.[13]

3. Instrumentation of the System

3.1. *The F.D.S. Generator*

The generation of f.d.s. as described by eqn. (3) involves two operations:

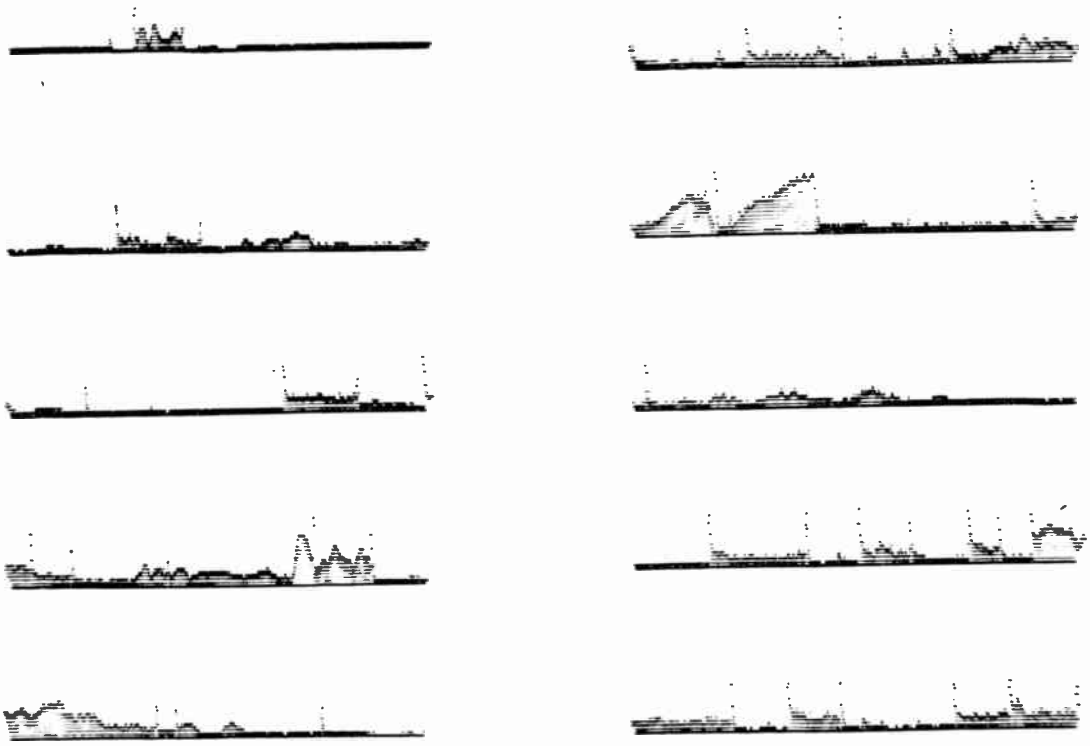
(i) producing a video signal delayed by one frame interval T, and

(ii) subtracting the delayed from the undelayed signal or vice versa.

Whereas the second operation is comparatively simple, the first one requires special components. It is obvious that to obtain a linear delay of 40 ms of a signal wave-form having 80 ns minimum rise-time, i.e. a delay-to-rise-time ratio of 500 000, represents an impracticable task using passive networks. Therefore only magnetic or electronic storage devices can be considered.

In the magnetic storage field analogue signal storage on magnetic tape appears feasible, but only if non-contacting heads are employed, because otherwise the tape surface and head wear would for present-day materials limit the operation to a very short time. Devices with non-contacting heads are being developed at present[14,15] and may be available soon.

Digital magnetic storage would involve the storing of 400 000 words of 5 to 6 bits each at the rate of 10^7 words per second. Since for the f.d.s. generating process only sequential word access is required, a high density magnetic drum may be considered. However, space multiplexing over say 50 tracks would be necessary, even if a longitudinal packing density of 1000 bits per inch can be achieved. Magnetic thin-film static stores are another alternative but, at the present state of the art, stores of the required capacity seem to be still difficult to construct, although steady and promising progress is being made in the automatic fabrication and micro-miniaturization of such devices and the associated circuits.



(a) Horizontal scale: 20 seconds per trace (500 frames).

(b) Vertical scale: 2·5% per step.

Fig. 4. Frame difference signal density recordings.

In the field of electronic storage devices, storage tubes are produced by many of the major electronic manufacturers. Most of these, being specially made for radar and other utilitarian applications, have still deficiencies which, however, seem to be not so much a matter of principle than of production techniques and specification. Recently, a new process for the storing of either wide band analogue or high bit-density digital signals was discovered and is in the stage of industrial development. This has been designated 'thermo-plastic' recording.[16] From all reports so far given in the literature, devices involving this method show promising characteristics for the storage of television signals.

The solution adopted here for the construction of the experimental f.d.s. generator represents a compromise between economy and purpose. Since the realization of a suitable storage device for full scale f.d.s.-coding operations will require an intensive engineering effort, this can only be justified if substantial savings in transmission channel capacity requirements can be achieved. On the other hand the data on f.d.s. densities so far available do not permit to make reliable estimates of the attainable results. It was therefore decided to concentrate in the first instance on the statistical analysis of the f.d.s. In an interlaced scanning system a strong correlation exists between the two fields in each frame. It was therefore considered permissible to suppress every second field, and thus reduce the number of elements for analysis from 400 000 to 200 000. Moreover, due to the blanking interval between each scanning line the active picture elements from which alone a frame difference signal can arise take up only about 80% of the video signal. Considering further that movement is generally concentrated in the centre of the picture, it was decided to limit the f.d.s. evaluation to the central 50% of the active picture area, i.e. to about 80 000 elements. (This involved suppressing 15% in linear dimension all around. The results would tend to be somewhat pessimistic, which should be acceptable.) By imposing these limitations it was expected that a technique of f.d.s. generation could be successfully used which was suggested to the author by H. R. Day (General Electric Research) in 1958 and was independently published by Teer.[17]

A standard vidicon television camera tube is employed in the method chosen for the delay, i.e. storage, and subtracting operation. In spite of the previous publication of this technique, it was found necessary to investigate the operating conditions and control of deficiencies of the vidicon for this purpose

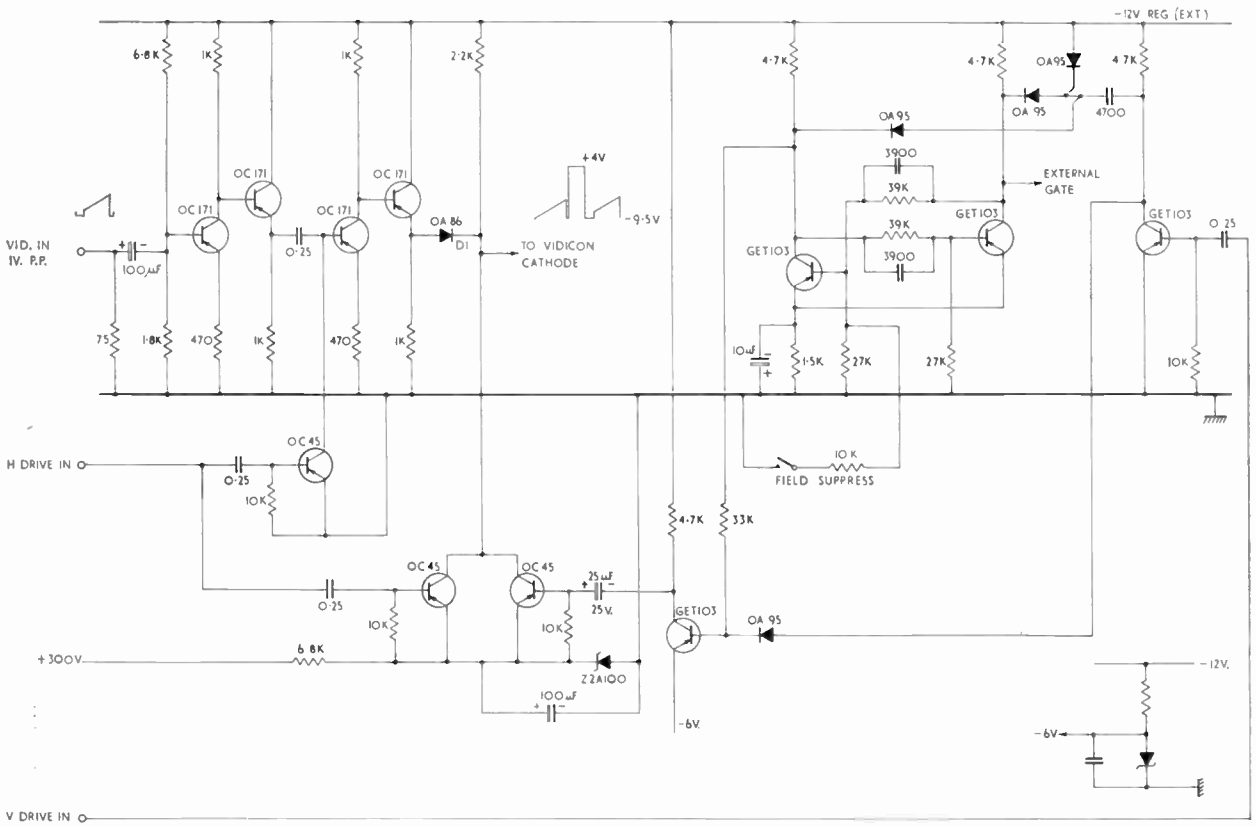


Fig. 5. Cathode modulating amplifier assembly.

more fully than initially expected. A co-worker of the author, J. B. Potter, has reported in detail about these investigations[18] so only the operational principles of using a vidicon as a storage device for television signals, and specifically for f.d.s. generation will be discussed here.

The main feature of this application of the vidicon is the fact that the cathode of the tube is modulated with the signal to be processed. Since in normal operation the scanning electron beam charges the vidicon target to cathode potential, cathode-modulation will cause a corresponding spatial potential distribution on the target. If now in the next scan the beam strikes the target again and the cathode potential is the same at the particular time, i.e. the particular element has the same amplitude, ideally no current will flow into or out of the target, thus no signal will be generated in the target load. If, however, the cathode is more negative at that instant, current will flow out of the target until that area is charged to the new potential and a signal will be generated which is proportional to the difference of the cathode potential, i.e. of the amplitude of the video signal, at the same time in two consecutive frames (f.d.s.). But, if the cathode were more positive no electrons could reach the target and thus no signal would be generated in spite of an existing difference. In order to overcome this difficulty the photo-conductive target is illuminated uniformly by an external light bias source. Due to the conductivity of the target which this source generates, the target potential will drift between consecutive hits by the electron beam towards the positive voltage applied to the signal electrode ('target-voltage'). If light bias and target voltage are chosen so that this potential drift of the target in volts is larger than the maximum amplitude of the signal applied to the cathode, the cathode can never become more positive than the target, and f.d.s. will be generated in positive and negative directions about a fixed direct voltage component caused by the light bias drift.

From this description of f.d.s. generation by means of a vidicon camera tube it is apparent that apart from the operational components of a normal television camera chain, the only additional equipment is a suitable video amplifier for modulating the cathode of the vidicon with the video signal and an arrangement to apply a uniform light bias to the target area.

Although the experimental prototype of the device was especially assembled from industrially produced components, a studio type vidicon chain was used for the more general investigations reported by Potter,[18] and therefore only the specific circuits and arrangements required for the generation of frame difference signals will be discussed here. If the camera chain does not contain extensive metering facilities for the monitoring of all voltages and currents applied to and flowing in the vidicon, these must be provided because of the significant effect of these data on the frame difference signal.

An ordinary slide projector may be used for the applying of light bias. This offers the advantage of setting up electrical focus, scanning raster size, linearity and centring by means of suitable test slides. Raster linearity, beam landing error, light bias-, target- and focus-uniformity have been found to be the most critical operating parameters, in fact, much more critical than in normal camera operation. All these parameters may affect the uniformity of light bias drift, thus superimposing spurious waveform components on the zero axis of the f.d.s. These spurious components, apparent as 'shading' signals, are also proportional, but not necessarily linearly related to target voltage and light bias flux, i.e. to the amount of target potential drift between consecutive scans.[18] Keeping the potential drift small, however, would reduce the permissible peak amplitude of the cathode modulating signal. In consequence the peak amplitude of the target output signal would be reduced and for a given thermal noise voltage generated in the camera output amplifier, the signal to random noise ratio of the f.d.s. would become smaller.

Therefore, in order to minimize the detection threshold ε (eqn. (4a)) for the processing of the f.d.s., apart from striving for extreme precision of the deflection waveforms and system, and introducing compensating waveforms for beam landing error, tubes having most uniformly high target sensitivity are required. Moreover, the noise figure of the camera amplifier should be as low as possible. Although the reduction of the active picture area for analysis made it possible to select the best area of the target for minimum shading signal and also mitigated scan linearity requirements, more development work is planned to reduce the remaining deficiencies still further.

The transistorized cathode modulating amplifier, shown in Fig. 5, accepts video signals at standard level (1 V peak-peak) as input and supplies up to 3·5 V peak signal to the cathode. Apart from the actual amplifier, the second stage of which is clamped by horizontal-drive, the assembly contains also the facilities for beam-blanking. This is carried out by clamping the vidicon cathode to a positive voltage during horizontal-drive, vertical-drive and during one whole field period, the video signal being disconnected by D1 becoming non-conducting during beam-blanking. The driving signal for the field suppression action is derived from a flip-flop stage operating from vertical-drive signals. The same flip-flop also supplies one of the control signals for the gate by which the output signal of the f.d.s. generator is restricted during processing and for the reset-pulse timing of the sample counter.

The raw f.d.s. waveform obtained at the camera amplifier output has a peak-to-peak amplitude, equivalent to that of standard video signals, but is bi-directional relative to approximately 'mid-grey' level. This signal is then processed as described in the following section.

3.2. *F.D.S. Processing*

The processing equipment itself contains three sections, the signal-restricting gate, the slicing and sample generating section and distribution amplifiers (Fig. 6).

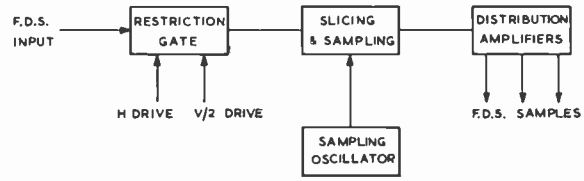


Fig. 6. Processing equipment block schematic.

The signal-restricting gate consists of a standard video gate using saturated transistors as series switches as described by Potter.[19] This gate having been constructed for general purpose laboratory applications has two complementary gating stages, so that it fulfils the function of a single-pole-double-throw switch. However, in this application only one of the input channels is used being switched on for those time intervals of each scanning line and frame, during which the f.d.s. is analysed.

Four delay multivibrators (DMV) serve for the generating of the gate-actuating waveform (Fig. 7). By making the time constant of these DMV's variable, the onset as well as the width of the gates can be varied. Thus the delay of DMV-H1 and DMV-V1 being triggered from H-drive and V/2-drive (from field suppression flip-flop in cathode modulating assembly) respectively determines the start of the gating waveforms in horizontal and vertical direction generated by DMV-H2 and DMV-V2, the delay of which in turn determines the duration of the gating waveforms. Provision is made in the two gates for

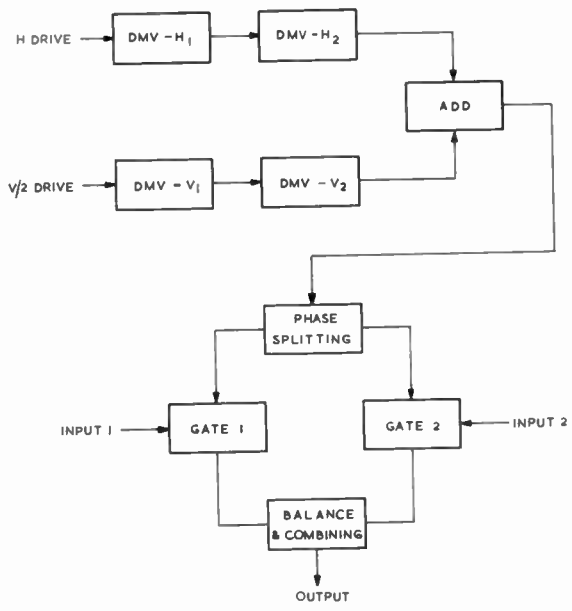


Fig. 7. Restriction gate.

balancing the output of the two signal paths, a feature which in this application is used to set the zero f.d.s. level for the subsequent slicing operation.

The non-linear processing of the f.d.s. functionally represented by Fig. 2 and eqn. (5) is carried out by means of the circuits shown in Fig. 8. In order to obtain a more defined threshold action the gated f.d.s. is first amplified. The two-stage amplifier, VT1 and VT2, serves at the same time as limiter. This dual action is achieved by using non-linear collector loads consisting of diode pairs (D1, D2 and D3, D4) connected back to back. These diodes, biased by D5 and D6 let the two stages have maximum gain only, when the diodes are open circuit, whereas as soon as the collector voltages exceed the bias voltages in one or the other direction, the collector load resistors are shunted by the low resistance of the diodes and the bias network. The bias voltages of the two diode pairs as well as that of the two diodes in each pair are set-off so that the linear swing of VT1 is about 1 V and that of VT2 about 1·5 V. Obviously for symmetrical limiting action the no-signal operating point of both stages must be in the centre of their linear range. This condition is achieved by the choice of the Zener diode D7 used for interstage coupling and by presetting the collector voltage of VT2. For monitoring the proper adjustment of this amplifier a test-point is provided at the emitter of VT3.

The bi-directionally limited f.d.s. is now further processed for the generating of the f.d.s. gated Nyquist samples. (The author is indebted to his co-worker J. V. Murphy for the design of these circuits.[20]) Inversely-identical limited f.d.s. signals are obtained at the collectors of VT4, VT5 and fed to the parallel transistor pair VT6, VT7. This pair is biased and current-switched by VT8 and D11 so that collector current can flow only during those portions of the signals at the two bases which lie in a negative direction relative to zero-f.d.s. level. This collector current is limited by the bias on the current source VT8, thus

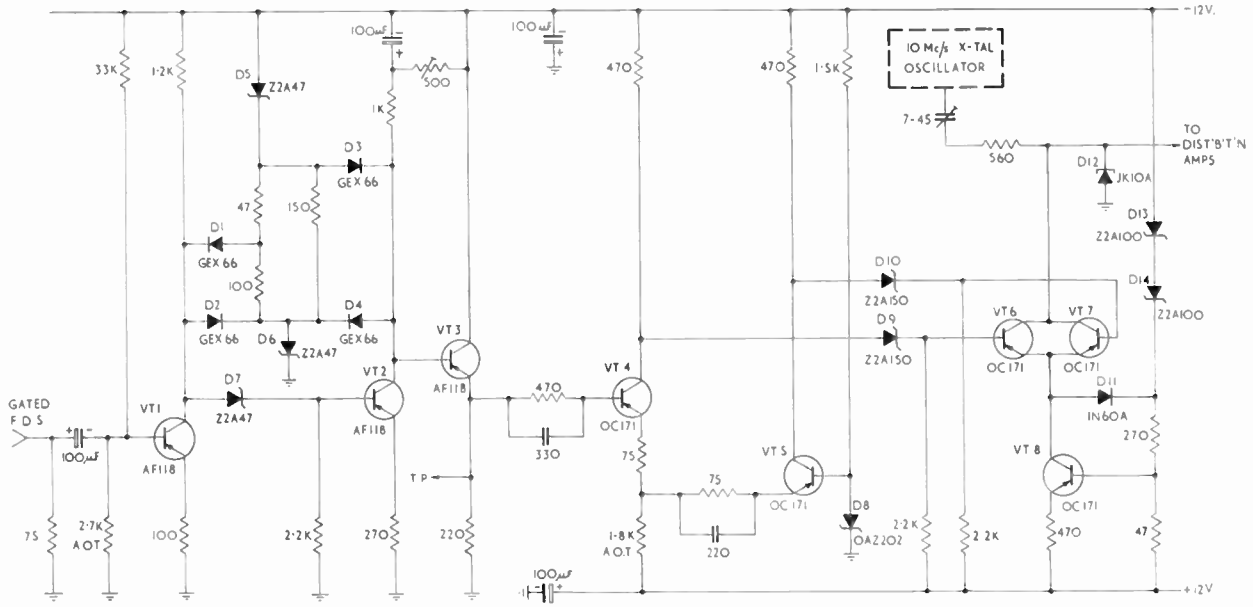


Fig. 8. Slicing and sampling circuit.

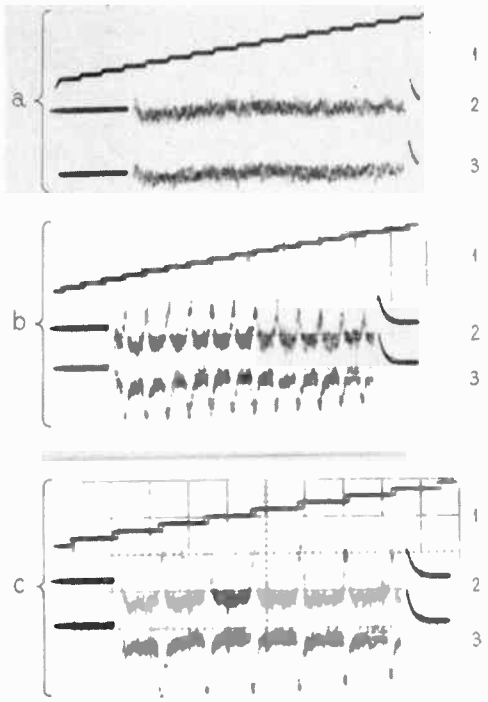


Fig. 9. Frame difference signal test waveforms. Group (a) is at the top, (b) central and (c) at the bottom. Each group consists of 3 traces, denoted 1, 2 and 3. Horizontal scale: 5 μs per major division. Vertical scale: traces 1, 1 V/major div., traces 2 and 3, 0·5 V/major division.

represents the absolute f.d.s. limited to ε' (Fig. 2) which acts as bias current on the tunnel diode D12, but is itself made not large enough to exceed the peak current (low voltage) point of the diode-characteristic. In parallel with the bias current a 'pumping' current supplied by a 10 Mc/s crystal oscillator is applied to the tunnel diode. With no f.d.s. current flowing this pumping current itself will not suffice to drive D12 beyond the peak current point, however, as soon as f.d.s. current starts flowing the two currents combined cause D12 to switch to the high voltage region beyond the valley point. By the biasing of VT6 and VT7 and by adjusting the coupling capacitor to the 10 Mc/s oscillator (pumping amplitude) the action of the tunnel diode switch is controlled so that during every cycle of the pumping waveform and in the presence of f.d.s. current the voltage across the tunnel diode is switched from a point below the peak current point to one beyond the valley point. Consequently 10 Mc/s pulses, i.e. Nyquist samples, are obtained across D12 during f.d.s., whereby the threshold ε in Fig. 2(a) depends for a given f.d.s. current on the amplitude of the cathode potential of D11.

The functional performance of this part of the processing equipment may best be shown by means of the test waveform records depicted in Fig. 9 and Fig. 10. For these tests the cathode of the vidicon is modulated with a staircase waveform of line scanning frequency having 20 and 10 amplitude steps respectively between 'black' and 'white'. The waveforms shown were recorded for a selected line period in a sequence of about 10 frames. In each of the three records, Fig. 9(a), (b), (c), trace 1 represents the cathode waveform, traces 2 and 3 the waveform obtained at the emitter of VT3 (test point) for alternatingly consecutive frames. Since for Fig. 9(a) the input waveform does not change in time no f.d.s. is generated and the record (Fig. 9(a) 2 and 3) shows only the amplifier noise. Movement is then artificially generated by means of the device shown schematically in Fig. 11.

The input signal is fed to the vidicon cathode by an electronic switch either directly or via a delay line of 0·5 μs. This switch is operated from a binary divider chain which divides the field frequency (V) by four and thus selects during alternating frame intervals either the direct or the delayed video signal. The back and forth movement of the staircase so generated which is just recognizable in Fig. 9(b,1) and 9(c,1) causes a f.d.s. to arise on every step of the staircase, its polarity changing in consecutive frames as apparent from Fig. 9(b) and 9(c) traces 2 and 3. Because of the

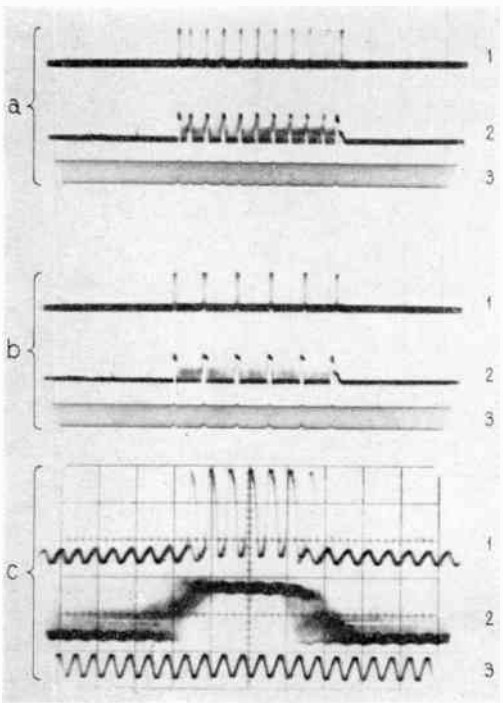


Fig. 10. Tunnel diode waveforms. Nomenclature as for Fig. 9. Horizontal scale: (a) and (b), 5 μs, (c) 0·2 μs per major division. Vertical scale: traces (1) 0·5 V; (a,2) and (b,2) 4 mA; (c,2) 2 mA; traces (3), 5 V for major division.

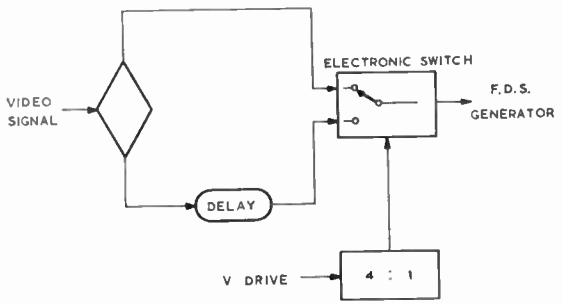


Fig. 11. Movement simulator.

compander action of the amplifier VT1 and VT2 the f.d.s. amplitudes for the 5% and 10% steps are practically identical.

Figure 10 now shows the corresponding waveforms associated with the tunnel diode D12. In each of the three records, a, b, c, trace 1 depicts the f.d.s.-gated Nyquist samples across D12, trace 2 the unipolar tunnel diode bias current and trace 3 the 10 Mc/s pumping voltage.

The records in Figs. 10(a) and (b) show the waveforms for the 20- and 10-step staircase respectively corresponding to Figs. 9(b) and (c), whereas in Fig. 10(c) the f.d.s. arising from a single step of the 20-step staircase is recorded on an expanded time-base. The blurriness of the edges of the tunnel diode current waveform in Fig. 10(c,2) and hence the intermittent appearance of the two Nyquist samples at the edge of the pulse 'packet' in Fig. 10(c,1) is due to the fact that the time-base was synchronized by the 10 Mc/s pumping oscillator which is not in synchronism with the line scanning frequency and thus the staircase steps. Figure 10 shows clearly the threshold and switching action of the pumped tunnel diode in as much as the bias current waveform (trace 2) still contains noise which is no longer apparent in the Nyquist samples (trace 1), i.e. the tunnel diode output waveform. The small-amplitude 10 Mc/s waveform existing in the latter during no-f.d.s. periods results from the voltage generated by the pumping current below the peak-current point of D12, but this is easily eliminated in the subsequent amplifier circuits, shown in Fig. 12.

These circuits serve the purpose of amplifying the Nyquist-sample pulses to the required amplitude, establishing correct polarity for subsequent operations and providing low impedance outputs for coaxial cable connections to other equipment. The pulses obtained from D12, having about 0·5 V peak amplitude and positive polarity, are first amplified by the npn transistor VT9 which is operated close to cut-off and thus suppresses the residual waveform component during no-f.d.s. This stage is followed by an emitter follower, VT10, supplying negative polarity pulses of 1·5 V amplitude to the f.d.s. density integrating counter. The same pulses are then inverted in VT11 and fed to two more emitter followers VT12 and VT13, which supply positive polarity pulses of 1 V amplitude to the display and (future) video sampling circuits. All outputs are loaded by 75 Ω terminations and also provide a source termination of close to 75 Ω.

3.3. *F.d.s.-density Recording*

The quantitative measurement of f.d.s.-densities and their variations in real time as defined by eqn. (7) and described in Section 2 implies the counting of the

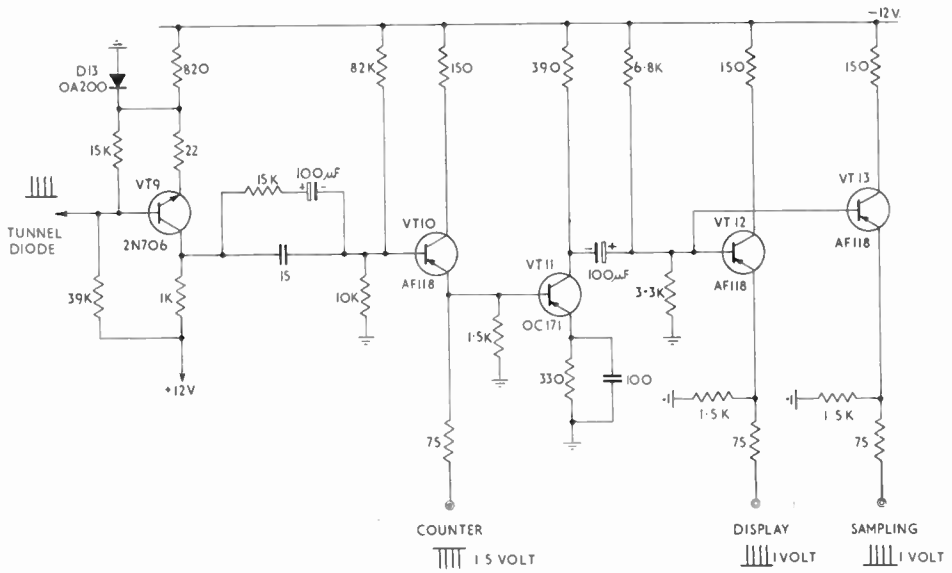


Fig. 12.
Pulse distribution amplifiers.

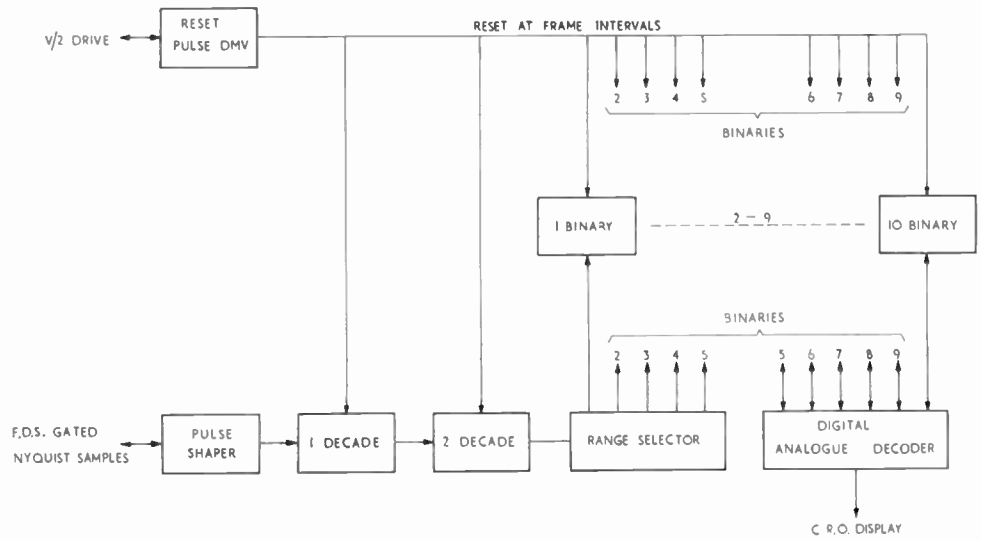


Fig. 13. Frame difference signal density recording.

f.d.s.-gated Nyquist-samples obtained at the output of the processing equipment. For this purpose a high-speed decade counter was developed in the laboratory. This counter decade consists of a ring of five binary stages sequentially switched as described by Tarczy-Hornoch.[21] In order to realize the high-speed requirements a non-saturating current steered flip-flop circuit[21,22] was chosen for the individual stages and adapted for locally produced semi-conductors. It would be beyond the scope of this paper to discuss the details of the actual circuit, but using OC171 transistors throughout, a double-pulse resolution of 25 ns was attained for random input without extreme circuit refinement, equivalent to a periodic input of 40 Mc/s.

As shown in Fig. 13, two such decades follow a pulse shaping circuit (a high-speed blocking oscillator), the carry-pulses of the second decade feeding into a chain of ten binary stages constructed of medium-speed commercially produced building blocks[23] (Philips type B8-920-00). The maximum count capacity of this counter is therefore 102 300, but in operation the maximum possible number of Nyquist samples n_F is set to only 80 000 per frame by adjustment of the restriction gate. Normally the last six digits of the 10 binaries are used for recording, thus the least significant digit represents 1600 samples or 2% of the total image area tested. It was, however, sometimes found desirable to achieve a better resolution and therefore a 'range selector' switch was incorporated which permits up to four binary stages being bridged. By this facility the value of the least significant digit may be reduced to 800, 400, 200 or 100 corresponding to a recording resolution of 1%, 0·5%, 0·25% and 0·125% respectively. It has hence to be noted that counter-overflow will occur when the count per frame exceeds 40 000, 20 000, 10 000 and 5000 samples, which could lead to ambiguities when unknown pictures are being analysed. The lower range settings of the selector are therefore primarily intended for calibration of the f.d.s. threshold setting and for noise density counts.

At the end of every frame all counter-stages are reset to zero. For this purpose a reset pulse is derived from the field-suppression flip-flop in the cathode modulation assembly (Fig. 5) by means of a DMV having a delay of 30 ms. Due to the field-suppression, samples are only generated and thus counted for the first 20 ms of each frame interval. The counter retains this count until reset by the reset pulse generated from the trailing edge of the DMV, i.e. 10 ms after the end of the active field period. This way the counter stays for at least 10 ms at the respective frame count and for another 10 ms at zero before commencing the counting of the next frame's f.d.s. samples.

So far only the photographic recording of count-analogue voltages has been instrumented (pending the supply of a high-speed paper tape punch for bulk-recording). For the generating of the count-analogue voltage a binary decoder network using weighted resistors[24] has been incorporated in the counter. In accordance with the principle of such a decoder, as the count progresses during each frame interval, the output voltage increases in quantum-steps representing the count-equivalent of the least significant digit (e.g. 1600). This causes the striated appearance of the photographic records (Fig. 4) and may readily serve as a scale for quantitative evaluation. If, for example, n_F was set to 80 000 this would produce 50 steps in the analogue voltage output and therefore each step would correspond to 2% f.d.s. density.

4. Discussion

It is well known to those familiar with the operation of television cameras that a great many inter-dependent variables affect the performance of these devices. This is even more disturbing, if such a device is being used, not just as a 'picture signal generator', but as an objective measuring device from which reproducible results are expected within definable confidence-limits.

In the system described here, the validity of the measured values of f.d.s.-density depends primarily on the threshold ε which, when exceeded by the raw f.d.s. (Fig. 2) leads to the generation of f.d.s.-gated 10 Mc/s sampling pulses. Although in this context these pulses are only used for measuring the f.d.s.-density by the process of numerical integration expressed by eqn. (7), in an actual f.d.s. coder they will also serve to select those samples of the original video signal which have thus been found to be different in consecutive frames and are therefore required to be coded for transmission.[10]

If ε is too high, low-amplitude differences will not be recognized by the processor and consequently not be counted and coded, if it is too low a substantial number of 'noise pulses' will be identified as f.d.s. At least a value of 3% appears desirable for ε where differences in large areas are involved, corresponding to an equivalent 5-bit quantization of the original signal. Figures 9 and 10 indicate that this threshold value can be attained when random noise is the only determining factor. However, shading signals superimposed on the random noise and caused by non-uniformity of the vidicon target, imperfect beam alignment, deflection non-linearities and focus modulation may affect the attainable threshold setting substantially.

In the first instance the threshold setting is indicated by the ratio of the noise count (no signal input) to the f.d.s. density count for a predictable and thus objectively measurable f.d.s. test input. Apart from continuing investigations directed towards minimizing the shading signals by selecting the most suitable vidicon type and more sophisticated ancillary circuits, as well as developing a camera head amplifier having a lower noise figure, efforts are being made to devise objective test signals and methods for determining the threshold setting reproducibly and for all conditions expected in normal television signals. The test staircase depicted in Figs. 9 and 10 is only the first step in this direction. But the fact that the areas covered by the frame difference signals thus obtained are comparatively small introduces a considerable uncertainty and a number of other additional test signals are being investigated presently.

Therefore, at this stage, the records shown in Figs. 3 and 4 should only be considered as qualitative examples. However, it is thought that even so they are sufficiently unique to illustrate the achievements to date. The three screen photographs (Figs. 3(a), (b) and (c)) depict in this order a lady in conversation (camera stationary), two men (camera panning) and a group of riders on trotting horses (camera stationary). These particular three pictures were chosen from a number of samples recently recorded because their content is easily recognized and also because they are quite typical for actual television material.

The f.d.s.-density records (Figs. 4(a) and (b)), are taken from an actual television programme containing significantly more action than many others. Each of the five traces in the two figures represents approximately 20 seconds of record, i.e. a total of 5000 frames for the 10 traces, each frame count being taken over an area of about 64 000 elements.

Scene changes are clearly recognizable by the distinct spikes or abrupt changes in the structure of the record. Panning scenes are typically apparent as high f.d.s. density extending often over five and more seconds, as in the second trace of Fig. 4(a) and in the fourth and fifth of Fig. 4(b). Areas showing low f.d.s. density of rapidly changing magnitude, as in Fig. 4(a) on the fourth and fifth trace, are generally obtained for quite violent movements within the picture (camera stationary), whereas the uniformly low count results from typical 'conversation' scenes with no or little movement. The noise count for all these records was about one per cent and each quantum step indicated by the horizontal striations represents $2\frac{1}{2}$% of the total area, i.e. a count of 1600.

These records demonstrate very significantly the different modes by which the frame difference signals may arise and also show that a given mode may persist at least over one interval between scene changes, which in the average is of the order of five seconds. Moreover, when observing a number of different programmes, one can almost recognize from the dynamic structure of the f.d.s.-density record the 'hand-writing' of the individual producer.[25]

It is this non-stationarity of the f.d.s. which causes the main difficulty and the disappointingly low coding efficiency in applying purely statistical coding methods.[4] The solution to this problem is expected to be found in the approach indicated in the Introduction in which the coding procedure is adaptively controlled by the dynamics of the input signal and into which the dynamic response of the human sense of vision is projected. So far this approach could only be developed conceptually[9,10] but the recording real-time data in large quantities made possible by the construction of the described experimental system, shall provide a quantitative basis for the testing of what so far had to be only a hypothesis.

5. Conclusion

In this paper a system has been described in which a vidicon camera tube is used for the generating of television frame difference signals. The raw f.d.s. is processed so that 10 Mc/s pulses (Nyquist samples) are generated, whenever an amplitude difference exceeding a defined threshold occurs between identical picture elements in two consecutive frames. Counting these pulses over each frame interval a quantitative record of the f.d.s. density and its fluctuation in real time is obtained. This will facilitate the statistical analysis of data thus derived from actual television programme material and will form a quantitative basis for the testing of f.d.s. coding hypotheses.

6. Acknowledgments

Apart from the contributions by Messrs. Potter and Murphy, already mentioned, the author wishes to acknowledge the assistance in the construction of the mechanical and circuit assemblies provided by Messrs. MacLachlan, Horrocks and Lieske.

For permission to publish this paper the author is indebted to the Engineer-in-Chief of the Postmaster-General's Department.

7. References

1. R. D. Kell, British Patent No. 341811, 1929.
2. F. Schroeter, "Speicherempfang und Differenzbild im Fernsehen", *Archiv der Elekt. Ubertr.*, 7, p. 63, 1953.
3. E. R. Kretzmer, "Statistics of television signals", *Bell Syst. Tech. J.*, **31**, p. 751, 1952.
4. Y. Taki, N. Tanaka and A. Yamazaki, "Statistical characteristics of television signals and estimation of a limit of bandwidth reduction", *Tech. J. Japanese Broadcasting Corp.*, **12**, p. 26, 1960.
5. B. M. Oliver, "Efficient coding", *Bell Syst. Tech. J.*, **31**, p. 724, 1952.
6. R. M. Fano, "Transmission of Information" (John Wiley, New York, 1961).
7. A. J. Seyler and Z. L. Budrikis, "Measurements of temporal adaptation to spatial detail vision", *Nature*, **184**, p. 1215, 1959.
8. A. J. Seyler and Z. L. Budrikis, "Measurements of Human Observer Response to Changes in the Content of Visual Displays", Comm. of Australia, P.M.G.'s Department, Research Laboratory, Report No. 5248, 1960.
9. A. J. Seyler, "Visual communication and psycho-physics of vision", *Proc. Instn Radio Engrs Aust.*, **23**, p. 291, 1962.
10. A. J. Seyler, "The coding of visual signals to reduce channel-capacity requirements", *Proc. Instn Elect. Engrs*, **109C**, p. 676, 1962. (I.E.E. Monograph No. 535E, July 1962.)
11. Y. Taki, N. Tanaka and K. Matsumara, "Television signal frame correlation and amplitude distribution measurements", *Tech. J. Japanese Broadcasting Corp.*, **11**, p. 5, 1959.
12. J. R. Ragazzini and G. F. Franklin, "Sampled-data Control System" (McGraw-Hill, New York, 1958).
13. Colin Cherry, "The bandwidth compression of television and facsimile", *J. Telev. Soc.*, **10**, No. 2, p. 40, 1962.
14. J. H. Wessels, "A magnetic wheel store for recording television signals", *Philips Tech. Rev.*, **22**, p. 1, 1960.
15. H. G. Walter, "Aufzeichnung und Wiedergabe von Standbildern mit dem Folienspeicher", *Elektronische Rundsch.*, **16**, p. 97, 1962.
16. W. E. Glenn, "Thermoplastic recording", *J. Soc. Mot. Pict. Telev. Engrs*, **69**, p. 577, 1960.
17. K. Teer, "Investigations into redundancy and possible bandwidth compression in television transmission", *Philips Res. Report*, **15**, p. 30, 1960.
18. J. B. Potter, "On the use of the vidicon camera tube as a video storage device", *Proc. Instn Radio Engrs Aust.*, **24**, p. 855, December 1963.
19. J. B. Potter, "A video gating system using a transistor as the active element", *Proc. Instn Radio Engrs Aust.*, **24**, p. 422, 1963.
20. J. V. Murphy, "A Tunnel Diode Pulse Threshold Gate", Comm. of Australia, P.M.G.'s Department, Research Laboratory Report No. 5766.
21. Z. Tarczy-Hornoch, "Five-binary counting technique makes faster decimal counting units", *Electronic Design*, **9**, p. 34, 1961.
22. H. S. Yourke, "Millimicrosecond transistor current switching circuits", I.R.E.–A.I.E.E. Conference on Transistor and Solid-state Circuits, Philadelphia, 1957.
23. A. C. Denne and N. A. Steadson, "Applications of Philips circuit blocks", *Miniwatt Digest*, **1**, p. 191, 1962.
24. A. K. Susskind, "Notes on Analog-digital Conversion Techniques" (The Technology Press, M.I.T., Cambridge, 1957).
25. A. J. Seyler, "Real time recording of television frame difference areas", *Proc. Inst. Elect. Electronics Engrs*, **51**, p. 478, 1963.

Manuscript first received by the Institution of Radio Engineers Australia on 29th January 1963 and in revised form on 19th June 1963. (*Paper No. 957/T30*)

Report of the Annual General Meeting of the Institution†

The Institution's Third Annual Meeting since incorporation by Royal Charter was held in London on 9th December, 1964, and was attended by sixty Corporate Members and a number of non-Corporate Members. The chair was taken by Mr. J. L. Thompson (the retiring President).

After the formal opening business of the meeting, the President moved the adoption of the Annual Report of the Council for the year ended 31st March, 1964. (Published in the *Proceedings* of the I.E.R.E. October-December, 1964‡). After expressing his personal appreciation of the work of his fellow members serving on the Council and the Standing, Group and Local Section Committees, and to the Secretary and his staff, Mr. Thompson referred to two events of great significance which had occurred since the end of the period covered by the Report. The Engineering Institution's Joint Council had submitted a petition to Her Majesty in Council for the granting of a Charter and he felt that it was encouraging to note that the aims and objects referred to in the petition had the unanimous approval of the thirteen chartered engineering institutions which formed the Council. The formal incorporation in July of the National Electronics Research Council was another event of tremendous importance to radio and electronic engineers.

The President continued:

"The Annual Report shows that the Council is well aware of its responsibilities to the whole of the Engineering Industry, not only in regard to the training of professional engineers but in creating better facilities for training technicians. There has now been formed a Society of Electronic and Radio Technicians. This is a natural result of the work of the Radio Trades Examination Board which our Institution helped to found some twenty-five years ago."

The President concluded by saying that he was very pleased to call upon the first Chairman of the Indian Division of the Institution, Major General B. D. Kapur, to support the motion for the adoption of the Annual Report.

After congratulating the President and Council on presenting such an interesting Annual Report, which he felt was evidence of the progress made by the Institution, General Kapur continued:

"I am of course pleased that in this report particular reference has been made to the development of Institution activities in India. Without doubt our own Institution has made a tremendous contribution to Commonwealth understanding by the way in which it is approaching the problems of Commonwealth members who live outside Great Britain. We in India no longer feel that we are just members in correspondence only.

The Report was approved unanimously.

The President then called upon Mr. G. A. Taylor, Honorary Treasurer, to move the adoption of the Accounts for the year ended 31st March, 1964.

Mr. Taylor said that he would draw attention to just two points of major significance, namely the increased expenditure on administration and on publications. The most practical method of reducing publication costs would be to increase advertising revenue.

The balance sheet showed that the building reserve had now been exhausted in carrying out many improvements at 8 and 9 Bedford Square, and the Finance Committee hoped that future contributions to the Institution's Building Reserve Fund would enable further extensions and improvements to be made to the premises. The accounts were approved unanimously.

The next item concerned the election of Officers and Council for 1964-65. As there had been no opposing nominations to those made by the Council, the President declared the following elected:

The President: Colonel G. W. Raby, C.B.E.

Vice Presidents: A. A. Dyson, O.B.E., I. Maddock, O.B.E., R. H. Garner, Major General B. D. Kapur, H. F. Schwarz and Professor E. Williams.

Ordinary Members of Council: T. A. Cross, Professor W. A. Gambling and M. James (Members); and R. J. Cox, Rear Admiral C. R. Darlington and N. L. Garlick (Associate Members).

Honorary Treasurer: G. A. Taylor.

The Business of the Annual General Meeting concluded with the presentation of Premiums and Awards for outstanding papers, and Examination Prizes. (See *The Radio and Electronic Engineer*, November 1964, page 294, and the Annual Report of the Examinations Committee.)

After the close of the Annual General Meeting Colonel G. W. Raby gave his Inaugural Address "National Prosperity and the Engineer" (published in the January 1965 issue of *The Radio and Electronic Engineer*).

† A full report of the Annual General Meeting will appear in the January-February issue of the *Proceedings*.

‡ The *Proceedings* is sent only to members in Great Britain. Members resident overseas may obtain a copy of this issue containing the Annual Report free of charge on application to the Secretary.

U.D.C. 518.5: 621.372.57

A Computer Program for Analysing Networks containing Three-terminal Active Devices characterized by their Two-port Parameters

By

L. R. FAIRBROTHER
(*Graduate*)†

AND

H. G. BASSETT, B.Sc.(Eng.)‡

Summary: The paper describes a general-purpose computer program which may be used to calculate the performance of a wide variety of 2-port networks. The circuit to be analysed may contain 3-terminal active devices (such as transistors) provided their 2-port parameters are known, and there are no restrictions (apart from an upper limit to the size of the circuit) on the manner in which the resistors, capacitors, inductors and 3-terminal devices may be joined together. Thus the performance of amplifiers having multi-loop feedback paths can be analysed. The program is written in Elliott 803 Autocode, and can analyse a large feedback amplifier in a few minutes.

1. Introduction

The circuit behaviour of the transistor is so complicated at high frequencies that it is often impracticable to devise equivalent circuits which are sufficiently accurate to be used in circuit analysis. A need exists for computational methods which can use the measured 2-port parameters of transistors in the analysis of such circuits as feedback amplifiers. Several workers[1,2,3] have described computer programs suitable for general analysis of circuits, and the present-day low cost of computer time justifies the use of digital computers to reduce the need for empirical experiments in circuit development.

The program described in this paper facilitates the analysis of 2-port networks consisting of R, L, and C elements together with 3-terminal devices (including active elements such as transistors). It is written in Elliott 803 Autocode and can accommodate up to 30 nodes in the circuit to be analysed. The form of the input data has been made as simple as possible; two data tapes are required, one giving the frequencies and the *y*-parameters of the 3-terminal devices, and the other containing details of the R, L and C elements and their locations. The program is broadly similar to the NAPANS[1] program used by the Bell Telephone Laboratories, but is restricted to 2-port networks, is arranged to accept data in a very simple form, and can deal with a larger number of nodes.

2. Node-datum Analysis

The program is based on node-datum analysis.[4] Figure 1 shows a general network having n terminals in addition to an earth terminal, with the current and voltage conventions indicated (all voltages are expressed relative to earth potential). The usual linear network equations can then be written as follows:

$$I_1 = Y_{11}V_1 + Y_{12}V_2 + \ldots Y_{1n}V_n$$
$$I_2 = Y_{21}V_1 + Y_{22}V_2 + \ldots Y_{2n}V_n$$
$$\vdots$$
$$I_n = Y_{n1}V_1 + Y_{n2}V_2 + \ldots Y_{nn}V_n$$

leading to the array:

$$\begin{matrix} Y_{11}Y_{12}\ldots Y_{1n} \\ Y_{21}Y_{22}\ldots Y_{2n} \\ \vdots \\ Y_{n1}Y_{n2}\ldots Y_{nn} \end{matrix}$$

which is known as the nodal admittance matrix for the network.

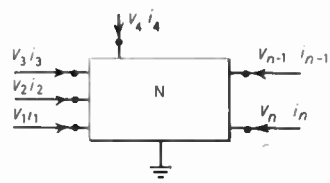


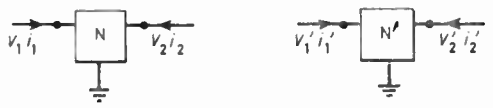
Fig. 1. General network with n nodes.

When two networks with n terminals are paralleled by joining corresponding terminals together, each term of the admittance matrix for the composite network is the sum of the corresponding terms of the admittance matrices of the two original networks. This addition follows because the new terminal currents are the sums of those of the two original

† Formerly at G.P.O. Research Station; now with International Computers and Tabulators Ltd.

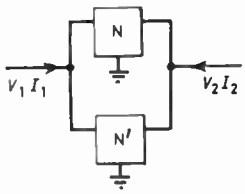
‡ G.P.O. Research Station, Dollis Hill, London, N.W.2.

networks, as is illustrated in Fig. 2 for a pair of 3-terminal networks, having each two nodes and an earth terminal.



$$i_1 = Y_{11}V_1 + Y_{12}V_2 \qquad i_1' = Y_{11}'V_1 + Y_{12}'V_2$$
$$i_2 = Y_{21}V_1 + Y_{22}V_2 \qquad i_2' = Y_{21}'V_1 + Y_{22}'V_2$$
$$Y\text{-matrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \qquad Y\text{-matrix} = \begin{bmatrix} Y_{11}' & Y_{12}' \\ Y_{21}' & Y_{22}' \end{bmatrix}$$

(i) Before paralleling



$$I_1 = i_1 + i_1' = (Y_{11} + Y_{11}')V_1 + (Y_{12} + Y_{12}')V_2$$
$$I_2 = i_2 + i_2' = (Y_{21} + Y_{21}')V_1 + (Y_{22} + Y_{22}')V_2$$
$$Y\text{-matrix} = \begin{bmatrix} Y_{11} + Y_{11}' & Y_{12} + Y_{12}' \\ Y_{21} + Y_{21}' & Y_{22} + Y_{22}' \end{bmatrix}$$

(ii) After paralleling

Fig. 2. Illustrating paralleling of two 3-terminal networks.

Thus we can build up the admittance matrix of a circuit of any desired complexity by allocating a set of nodes, writing down the separate indefinite admittance matrices[4] of all the components, including the active devices, and adding them together (an example is shown in Fig. 3). The computer can carry out this process automatically, so we need only provide data consisting of the component values and the nodes between which they are joined. An active device (such as a transistor) can be completely specified by its y-parameters and the three nodes to which it is connected.

Having obtained the admittance matrix for the complete network, we usually need to derive overall 2-port transmission performance; to do so we must make zero all the node currents except those at the input and output terminals. All voltages except those at the input and output are then eliminated from the nodal equations leaving two equations with four admittances which form the admittance matrix of the complete circuit, regarded as a 2-port. The computer will eliminate the unwanted nodes automatically by using a suitable procedure; pivotal condensation[5] was chosen because it is rapid in execution and occupies little storage space.

Thus, consider a set of three simultaneous equations relating the currents and voltages in a system with three nodes:

$$I_1 = Y_{11}V_1 + Y_{12}V_2 + Y_{13}V_3$$
$$I_2 = Y_{21}V_1 + Y_{22}V_2 + Y_{23}V_3$$
$$I_3 = Y_{31}V_1 + Y_{32}V_2 + Y_{33}V_3$$

To reduce this system to one with only two nodes, one node is made inaccessible by setting the current flowing into it equal to zero. If we choose to eliminate node 3 then

$$I_3 = 0 = Y_{31}V_1 + Y_{32}V_2 + Y_{33}V_3$$

Therefore

$$V_3 = (Y_{31}/Y_{33})V_1 - (Y_{32}/Y_{33})V_2$$

Substituting for V_3 in the equations for I_1 and I_2 gives a pair of equations representing the 2-node system.

$$I_1 = (Y_{11} - Y_{13}Y_{31}/Y_{33})V_1 + (Y_{12} - Y_{13}Y_{32}/Y_{33})V_2$$
$$I_2 = (Y_{21} - Y_{23}Y_{31}/Y_{33})V_1 + (Y_{22} - Y_{23}Y_{32}/Y_{33})V_2$$

The pivot point in this reduction is Y_{33}; principal diagonal terms are used as pivots because they cannot normally become zero. The process may be applied to a set of m simultaneous equations. The Y-parameters will form a square matrix of order m which is to be reduced to order $(m-1)$ by taking Y_{mm} as the pivot point. If the row suffix is denoted by i and the column suffix by k then the elements of the new matrix are given by

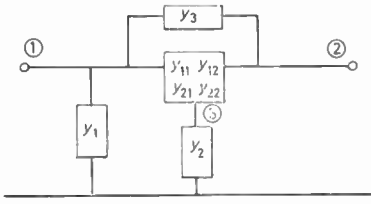
$$Y_{ik}' = Y_{ik} - (Y_{im}Y_{mk})/Y_{mm}$$

where

$$k = 1, 2, 3 \ldots (m-1)$$

for

$$i = 1, 2, 3 \ldots (m-1)$$



(i) Circuit configuration

	v_1	v_2	v_3
i_1	$y_1 + y_3 + y_{11}$	$-y_3 + y_{12}$	$-y_{11} - y_{12}$
i_2	$-y_3 + y_{21}$	$y_3 + y_{22}$	$-y_{21} - y_{22}$
i_3	$-y_{11} - y_{21}$	$-y_{12} - y_{22}$	$\Sigma y + y_2$

$$\Sigma y = y_{11} + y_{12} + y_{21} + y_{22}$$

(ii) Third-order admittance matrix

Fig. 3. An example of the formation of the admittance matrix for a complete circuit.

The reduction is easy to program. If the computer is programmed to start at Y_{11} and work systematically through the matrix row by row up to Y_{pp} (where $p = (m-1)$) then each new element Y'_{ik} may overwrite the old element Y_{ik} since that element is not required again. The value of m is then set equal to that of p and the process repeated until $m = 2$.

The basic instructions for this section of the program are:

```
→ p = m−1
  Calculate 1/Ymm
  Vary i = 1(1)p
→   Calculate Yim(1/Ymm)
    Vary k = 1(1)p
→     Calculate Y'ik = Yik − Ymk . Yim(1/Ymm)
      Repeat k
    Repeat i
    m = p
    Jump back to first line unless m = 2
```

3. Accuracy of Results

The equation used in the process of pivotal condensation shows how inaccuracies may arise in the computation. Under some conditions, for example at extremely low or high frequencies or at resonant frequencies, the admittance of some elements in the matrix can become very large compared with other elements in the matrix; thus the matrix tends to become degenerate.[4] Consider the situation where an inductance whose admittance is Y is connected between nodes 4 and 5 of a 5-node network, and there are resistors of admittance y between node 4 and earth and node 5 and earth.

Thus $Y_{44} = Y_{55} = Y+y$ and $Y_{45} = Y_{54} = -Y$. During the reduction process the equation,

$$Y'_{44} = Y_{44} - Y_{45} . Y_{54}/Y_{55}$$

i.e.

$$Y'_{44} = Y+y-Y . Y/(Y+y)$$

is evaluated.

As zero frequency is approached Y can become very large compared with y and Y'_{44} is then the difference between two large quantities. The accuracy of the Elliott 803 computer when using floating point arithmetic is nine decimal places, so that accuracy will be lost as the ratio Y/y approaches 10^9. If this ratio exceeds 10^9, y will be neglected in comparison with Y and the computer will evaluate Y'_{44} as zero or as a small quantity depending on rounding errors. The correct value of Y'_{44} as Y approaches infinity is $2y$, which may be seen by rewriting the equation for Y'_{44} as

$$Y'_{44} = (y^2/Y+2y)/(y/Y+1)$$

Although the original matrix may approach degeneracy as described above, the final second-order Y-matrix is nearly always well-conditioned. Consider, for instance, a ladder network in which an internal series element approaches zero in value. The nodal Y-matrix for the network approaches degeneracy, whereas the external electrical behaviour approaches that obtained by short-circuiting the element and suitably reducing the number of nodes. The final second-order matrix is now well-conditioned.

In practice the difficulty described occurs only very infrequently. When it does occur, the user of the program can readily avoid it by short-circuiting the offending element at the particular frequency concerned.

4. The Program

The program is capable of analysing a wide variety of circuits with the following restrictions (which could, if necessary, be altered by modifying the program):

(i) The maximum number of nodes is limited to 30.

(ii) The input and output ports chosen must have a common point, that is, they must form a 3-terminal network.

(iii) Any transformers contained in the circuit must be able to be represented as a 3-terminal network for which the y-parameters are known and are not excessively large.

A detailed flow diagram of the program is shown in Fig. 4. The basic steps in the program are as follows:

(i) Set up initial conditions.

(ii) Read the value of frequency. Clear two matrices $[A]$ and $[B]$ ready to store respectively the real and imaginary parts of the complex circuit admittances.

(iii) Process the list of two terminal components so as to build up the passive nodal admittance matrix in $[A]$ and $[B]$.

(iv) If the circuit contains active devices, add their y-parameters into the admittance matrices.

(v) Reduce the matrices by the method of pivotal condensation until they are of order 2.

(vi) Calculate from these second order matrices any desired circuit characteristics.

(vii) Repeat steps 2–6 inclusive until the circuit is analysed over the complete frequency range.

4.1. *Input Data*

For simplicity, the data required by the program have been divided into two sections 'A' and 'B', each section being punched on five-hole paper tape. Data tape 'A' is read by reader 1 while data tape 'B' (which must be in the form of an endless loop if it is required to analyse the circuit at more than one frequency) is read by reader 2.

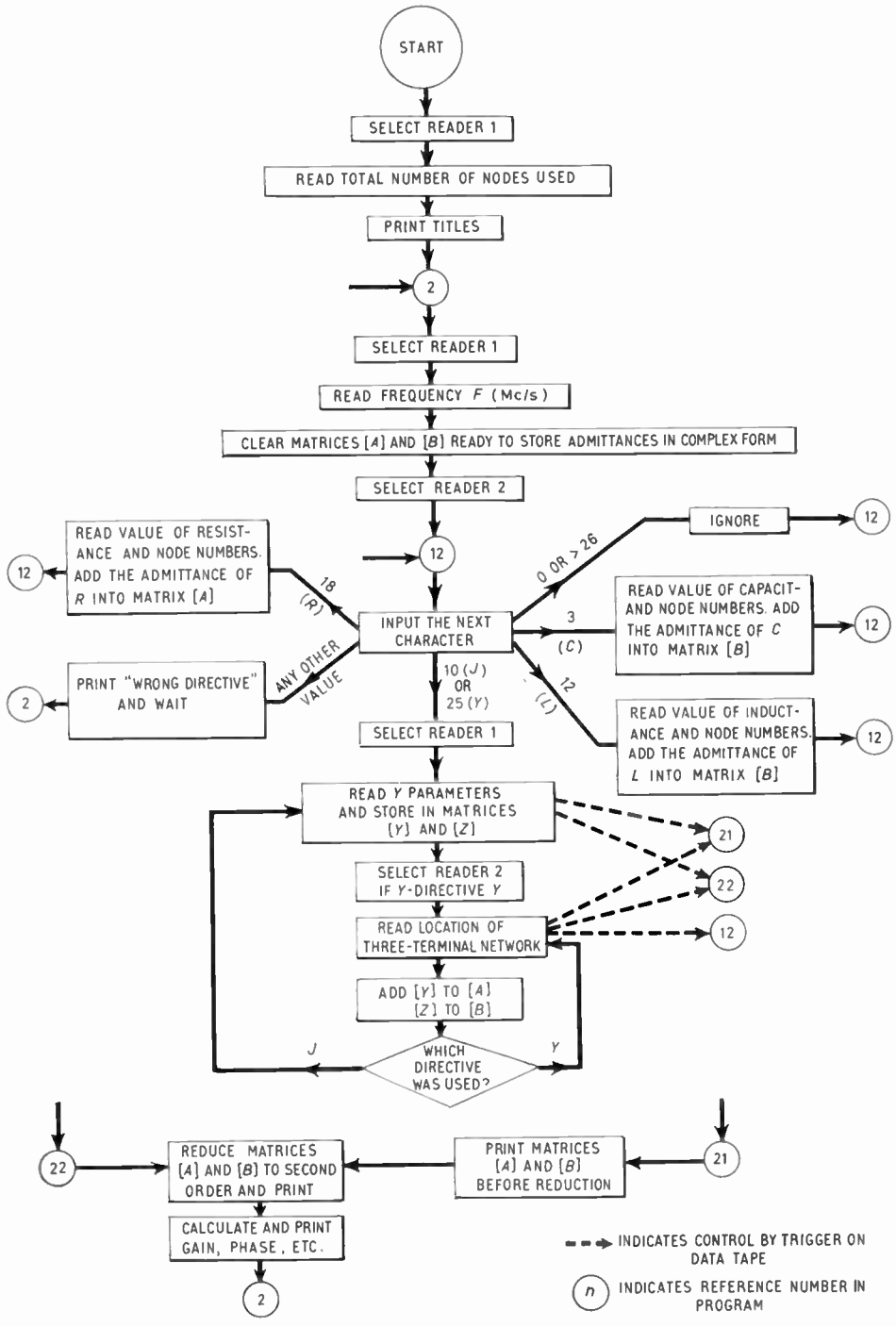


Fig. 4. Flow diagram for the program.

Data tape 'A' consists of:—

(i) The number of the highest node used.

(ii) A list of frequencies, each frequency being followed by a set of *y*-parameters (in complex form) for each type of active device present in the circuit.

(iii) A 'trigger' to stop or re-enter the program.

Data tape 'B' contains the following details of the circuit configuration:—

(i) A list, giving for each 2-terminal component in the circuit:

(*a*) a directive (R, L or C) indicating the type of component,

(*b*) the numerical value of the component, in ohms, microhenrys or microfarads as appropriate,

(*c*) the numbers of the two nodes between which the component is connected.

(ii) The directive *Y* followed by a list giving the three nodes for each 3-terminal network in the circuit.

(iii) A 'trigger' to enter the section of the program which carries out the matrix reduction.

If the circuit does not contain any 3-terminal networks then item (ii) above becomes simply the directive J. Provision is made to deal with more than one type of 3-terminal network if necessary.

The rules governing the allocation of nodes to the circuit are that node '1' should be assigned to the input terminal, node '2' to the output terminal and node '0' to the point common to both input and output terminals. Further nodes may be allocated in any desired order until every component is joined to two nodes, and every 3-terminal network is specified by three nodes. Care should be taken not to assign two nodes to the same point as this will cause the matrix to become degenerate.

4.2. *Output of Results*

Once the admittance matrix has been reduced to one of second order any desired circuit characteristic may be calculated. The program as written makes use of two output punches. On punch '2' the following quantities are printed out:

(i) The frequency in megacycles per second.

(ii) The second order admittance matrix in complex form.

(iii) The determinant of this matrix.

(iv) The voltage gain between nodes 1 and 2 in complex form.

Punch '1' has been used to accumulate the results as they are evaluated and gives the following quantities

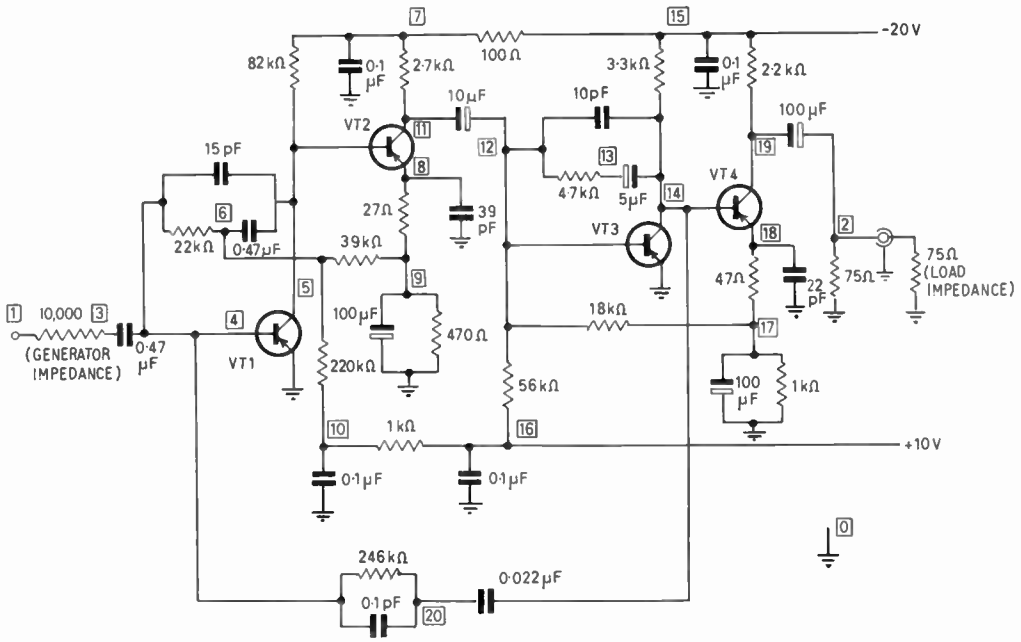


Fig. 5. Circuit diagram of a feedback amplifier showing allocation of nodes.

Data 'A'

```
20
 .10000/ 00
 .95246/-03   .39323/-04
-.61769/-08  -.29920/-05
 .38095/-01  -.81638/-04
 .12824/-07   .91256/-05
 .20000/ 00
 .95271/-03   .78645/-04
-.24707/-07  -.59839/-05
 .38095/-01  -.16327/-03
 .51294/-07   .18251/-04
.50000/ 00
 .95441/-03   .19660/-03
-.15441/-06  -.14958/-04
 .38091/-01  -.40815/-03
 .32056/-06   .45625/-04
 .10000/ 01
 .96050/-03   .39307/-03
-.61743/-06  -.29907/-04
 .38078/-01  -.81604/-03
 .12818/-05   .91229/-04
 .20000/ 01
 .98480/-03   .78513/-03
-.24666/-05  -.59738/-04
 .38028/-01  -.16300/-02
 .51208/-05   .18230/-03
 .50000/ 01
 .11532/-02   .19454/-02
-.15279/-04  -.14802/-03
 .37678/-01  -.40389/-02
 .31721/-04   .45300/-03
 .10000/ 02
 .17310/-02   .37716/-02
-.59244/-04  -.28697/-03
 .36479/-01  -.78301/-02
 .12300/-03   .88717/-03
 .20000/ 02
 .37267/-02   .67192/-02
-.21109/-03  -.51124/-03
 .32336/-01  -.13950/-01
 .43824/-03   .16442/-02
 .50000/ 02
 .10778/-01   .95190/-02
-.74762/-03  -.72427/-03
 .17696/-01  -.19762/-01
 .15521/-02   .29606/-02
 .10000/ 03
 .16380/-01   .74730/-02
-.11739/-02  -.56860/-03
 .60658/-02  -.15515/-01
 .24370/-02   .40944/-02
)
```

Data 'B'

```
R  10000  1  3
C  0.47  3  4
R  22000  4  6
C  0.47  5  6
C  0.000015  4  5
R  82000  5  7
R  2700  7  11
C  0.1  0  7
C  0.000039  0  8
R  27  8  9
R  470  0  9
C  100  0  9
R  39000  6  9
R  220000  6  10
C  0.1  0  10
C  10  11  12
R  3300  14  15
R  100  15  7
C  0.1  15  0
C  0.00001  14  12
C  5  14  13
R  4700  13  12
R  56000  12  16
R  1000  16  10
C  0.1  16  0
C  0.022  14  20
R  246000  20  4
C  0.0000001  20  4
R  2200  19  15
C  100  19  2
R  75  2  0
C  0.000022  18  0
R  47  18  17
R  18000  17  12
C  100  17  0
R  1000  17  0
R  75  0  2
Y
4  5  0
5  11  8
12  14  0
14  19  18
22C
```

Fig. 6. Input data.

FREQ MC/S	GAIN DB	PHASE SHIFT	INPUT IMPEDANCE		OUTPUT IMPEDANCE	
.100000	20.12	356.4	.100/ 05	-.500/ 00	.369/ 02	-.135/-01
.200000	20.12	352.8	.100/ 05	.410/ 01	.369/ 02	-.190/-01
.500000	20.17	341.9	.100/ 05	.139/ 02	.369/ 02	-.460/-01
1.00000	20.25	322.1	.101/ 05	.285/ 02	.369/ 02	-.100/ 00
2.00000	19.31	276.1	.101/ 05	.399/ 02	.370/ 02	-.292/ 00
5.00000	7.583	189.5	.101/ 05	-.122/ 02	.367/ 02	-.831/ 00
10.0000	-5.591	140.5	.101/ 05	-.239/ 02	.364/ 02	-.134/ 01
20.0000	-20.75	85.82	.101/ 05	-.272/ 02	.357/ 02	-.186/ 01
50.0000	-44.92	357.2	.100/ 05	-.205/ 02	.346/ 02	-.260/ 01
100.000	-63.69	278.6	.100/ 05	-.153/ 02	.339/ 02	-.383/ 01

Fig. 7. Output from punch 1.

in the form of a table:

(i) Frequency (in Mc/s).

(ii) Voltage gain (in dB) between nodes 1 and 2.

(iii) Phase shift (in degrees) between nodes 1 and 2.

(iv) Input and output impedances (in ohms).

The impedances are printed in complex form and the phase shift is given between 0 and 360 degrees without, however, any indication of the number of complete revolutions involved.

The time taken for the computation is very variable; it not only depends on the number of nodes present in the circuit (i.e. size of matrix), but also on the complexity of the circuit. For a two-stage amplifier containing 10 nodes the time of operation was 25 seconds per frequency value: for a four-stage amplifier with 20 nodes the time taken was one minute per frequency value.

5. An Example of the Use of the Program

The following example shows the application of the program to analyse a transistor amplifier[6] with several feedback loops (see Fig. 5). For the purpose of this example the four transistors have been assumed to possess an identical set of *y*-parameters, and the performance of the amplifier is therefore different from that described in reference 6.

Figure 5 shows the circuit diagram and the allocation of the nodes and since the overall performance is required nodes 1 and 2 have been assigned to the input and output respectively. Data 'A' and data 'B' are shown in Fig. 6. Eleven frequencies have been chosen to cover, on a logarithmic scale, the range 0·1–100 Mc/s. The output from punch 1 is shown in Fig. 7. Only the results from punch 2 for the first two frequencies are shown in Fig. 8. Finally, Fig. 9 shows a graph of the frequency response for the amplifier as plotted from Fig. 7.

```
FREQUENCY .10000/ 00

Y-MATRIX

 .99601/-04    .49827/-08
 .64617/-11   -.14400/-11
-.27434/ 00    .16940/-01
 .27123/-01    .99647/-05

DETERMINANT

 .27015/-05    .11271/-08

VOLTAGE GAIN

 .10114/ 02   -.62826/ 00


FREQUENCY .20000/ 00

Y-MATRIX

 .99598/-04   -.40498/-07
 .54153/-11   -.14392/-10
-.27295/ 00    .34110/-01
 .27121/-01    .14006/-04

DETERMINANT

 .27012/-05    .29251/-09

VOLTAGE GAIN

 .10063/ 02   -.12629/ 01
```

Fig. 8. Output from punch 2 (first two frequencies only).

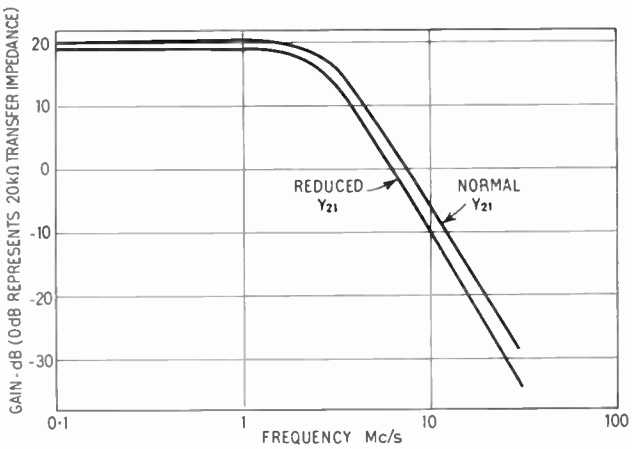


Fig. 9. Computed frequency response of amplifier, showing effect of reducing Y_{21} by 20% for all transistors.

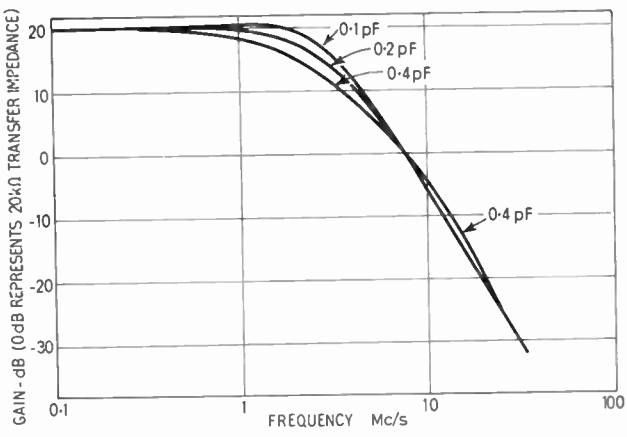


Fig. 10. Computed frequency response of amplifier, showing effect of varying feedback capacitance.

As an example of the use of the program for investigating tolerances, the computation was repeated with the Y_{21} parameters of all the transistors reduced by 20%. The resulting gain-frequency response is also shown in Fig. 9 for comparison. Figure 10 shows the effect of varying the feedback capacitance in the main feedback path.

6. Acknowledgments

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in the paper.

Acknowledgment is also due to Mr. A. J. Booth for assisting the work, and to Messrs. E. M. Cherry and D. E. Hooper for permission to make use of their amplifier configuration in the example.

7. References

1. O. P. Clark, "Design of transistor feedback amplifiers and automatic control circuits with the aid of a digital computer". *I.R.E. Intl Conv. Rec.*, **10**, Part 2, pp. 228–35, 1962.
2. F. H. Branin, "D-c and transient analysis of networks using a digital computer", *I.R.E. Intl Conv. Rec.*, **10**, Part 2, pp. 236–56, 1962.
3. G. H. Cohen and D. Platnick, "The design of transistor i.f. amplifiers using an I.B.M. 650 digital computer", *I.R.E. Trans. on Circuit Theory*, **CT-8**, No. 3, pp. 237–43, September 1961.
4. L. Weinberg, "Analysis of N-terminal networks", *Electrotechnology*, **69**, p. 64, 1962.
5. A. C. Aitken, "Determinants and Matrices". (University Mathematical Texts, 1942.)
6. E. M. Cherry and D. E. Hooper, "The design of wideband transistor feedback amplifiers", *Proc. Instn Elect. Engrs*, **110**, No. 2, pp. 375–89, February 1963 (I.E.E. Paper No. 4111E, 1963).

Manuscript received by the Institution on 15th May 1964. (Paper No. 958/C75.)

U.D.C. 621.385.644

Anode Structures for Cold-cathode High-power Magnetrons

By

YOSHITAKA IKEDA, Ph.D.†

AND

CHARLES SÜSSKIND, Ph.D.
(*Associate Member*)‡

Summary: The characteristics of anode structures suitable for cold-cathode high-power magnetrons have been investigated analytically and experimentally, with special emphasis on increasing the interaction and maximizing the area of coherent interaction with the electron beam at a given frequency. The structures analysed were designed for large mode separation, high interaction impedance, and easy coupling to the output circuit. They may be applicable to travelling-wave amplifiers; some circuit characteristics of the structures are therefore also included.

1. Introduction

The present contribution describes a project carried out as part of an investigation of cold-cathode, high-power, crossed-field devices that has occupied staff members and research assistants at this Laboratory over the past several years, notably under the direction of Prof. D. H. Sloan. Results of the investigation have been variously reported at scientific meetings in the U.S. and Japan and have achieved some circulation through technical reports such as those by Hoag,[1] Hartman,[2] Chamran,[3] Mercer[4] and (of particular relevance to the present contribution) Ikeda.[5,6]

Numerous interaction structures deemed suitable for the above-mentioned device type have been tested at this Laboratory. Among the considerations that enter into evaluating their relative merits are ease of fabrication, adequate mode separation, and ease of coupling to the load; the objective is to determine how (within these often conflicting requirements) the region of coherent interaction with the beam may be made as large as possible at a given frequency to maximize output power. At the same time, it is hoped that these investigations may throw new light on the understanding of the nature of crossed-field interaction.

As examples of conflicting requirements, consider that in a conventional magnetron the number of resonators is limited by the requirement of reasonable frequency separation of the modes. Cold-cathode interaction structures often have a large number of resonators. As the number of resonators increases, the mode separation of even the 'strapped' and 'rising-sun' configurations proves to be too low. On the other hand, some circuits (such as a corrugated cylindrical wall with interaction bars and coupling channels[2]) may have as much as 60% mode separation—but rather low interaction impedance for the TM_0 mode, since the electric field is minimum at the wall.

The new structures analysed in this report are free from this difficulty; however, they do suffer from the disadvantage of being susceptible to possible multipactor action along the magnetic field resulting from the r.f. electric field. This problem probably can be avoided by a careful design of the circuit.

The main assumption made in the present analysis is that the axial length of an individual cavity of the structure is much shorter than a free-space wavelength, so that the electric field inside the cavity is uniform in the axial direction. Moreover, it is assumed that since many bars orientated parallel to the axis are to be used in the structure, the capacitance between them and the cylindrical wall may be taken to be uniformly distributed along the cavity circumference. In most cases the capacitance between anode and cathode may be safely neglected.

M.K.S. units are used throughout and only the symbols given other than common meaning are explained in the text.

2. Proposed Anode Structures

The types of anode structures to be discussed are illustrated in Figs. 1, 2 and 3. The points at which the plate electrodes and bars are connected have been indicated by dots in the figures. The slow-wave structure of Fig. 1 is on the inside and is surrounded by a concentric cold cathode that depends on secondary emission for its electron current. The configuration shown in Fig. 2 is similar to that of Fig. 1, except that

† Department of Electrical Engineering, Yokohama National University, Japan; formerly at the Electronics Research Laboratory, University of California, Berkeley, U.S.A.

‡ Department of Electrical Engineering, University of California, Berkeley, U.S.A.

it is a conventional magnetron design, with the anode as the outer electrode. The TM_0 mode that is excited in each cavity makes the bars alternately positive and negative and thus provides the circumferential r.f. fields necessary for magnetron operation. There is a d.c. magnetic field in the direction of the z axis.† The configuration shown in Fig. 3 is suitable for a linear magnetron amplifier. A direct current flowing along the central structure could in principle produce a d.c. magnetic field, so that electrons would move in the axial direction under the action of both d.c. electric and magnetic fields. (Even though a large value of current would be required for this purpose, the corresponding direct voltage would be extremely small, so that power dissipation might be quite reasonable.)

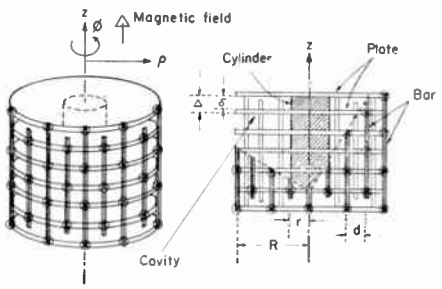


Fig. 1. Inverted magnetron-anode structure. Surrounding cathode cylinder is shown abbreviated. Bars act as pole pieces for magnetron.

The gap between the plates of the anode structures in Fig. 1 to 3 is denoted by δ, a parameter chosen so as to minimize the multipactor problem and to optimize the interaction impedance. In general δ is small and as a consequence the capacitance for a unit cavity is large, so that the frequency of the cavity is largely insensitive to the nature of electron-beam interaction or to the proximity of the cathode cylinder.

3. Resonant Frequencies for Individual Cavity

The structures proposed are particularly adaptable to multipole configuration up to several tens of poles (or bars). If a large number of bars are installed in a structure of Fig. 1 or 2, their effect on the cavity behaviour may be approximated by an equivalent uniform capacitance C_c per unit length in the circumferential direction. The edge effect of the plates may also be included in this equivalent capacitance.

† Such a structure could be also conceivably used for amplification, utilizing the wave travelling along the two groups of bars, which could be connected to external coaxial cables at both ends. In that case the interaction would be between the axial wave and the circumferential electron beam.

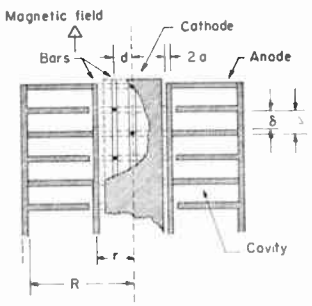


Fig. 2. Conventional structure for cylindrical magnetron.

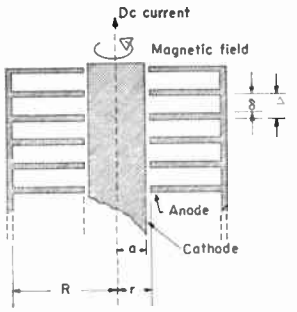


Fig. 3. Structure for linear magnetron.

Under the assumptions made in the previous section, Maxwell's equations are solved with the boundary condition at the open edges, where the capacitance C_c defines the ratio of the radial current to the r.f. voltage, giving the resonant frequencies with C_c as parameter.

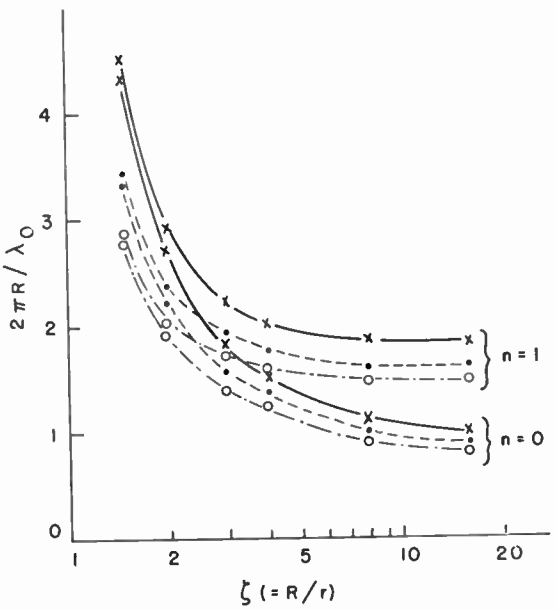


Fig. 4. Resonant frequencies for inverted magnetron structure. Values of $C_c\delta/\varepsilon_0 R$ denoted by × for 0, • for 0·1, ○ for 0·2.

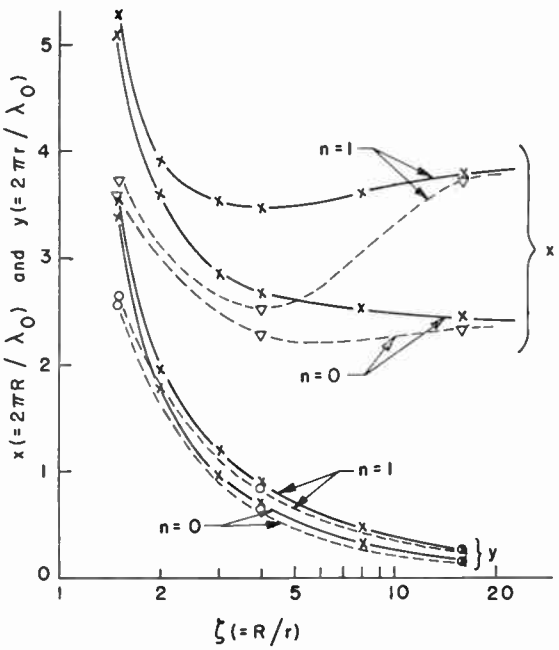


Fig. 5. Resonant frequencies for conventional magnetron structure. Values of $C_c\delta/\varepsilon_0 r$ denoted by × for 0, ⊙ for 0·2; $C_c\delta/\varepsilon_0 R = 0{\cdot}2$ denoted by ▽.

The frequencies of interest are the lowest possible, one corresponding to TM_{010} mode and the next one corresponding to TM_{011} mode; their calculated values as a function of the ratio of outer to inner radius $\zeta = R/r$ are shown in Figs. 4 and 5 for the two kinds of structures for unit cell. In these figures the two modes are denoted by $n = 0$ and $n = 1$; the ratio is taken as the abscissa. The degree of mode separation is also plotted in Fig. 6, where the parameter C_c is assumed to be zero. No appreciable change in mode separation is found even if C_c has an appreciable value (say, $C_c\delta/\varepsilon_0 R = 0{\cdot}2$, where ε_0 is the permittivity of free space). In Fig. 5 both $2\pi R/\lambda_0$ and $2\pi r/\lambda_0$ (where λ_0 is the free-space wavelength) are plotted for the conventional magnetron structure because of convenience in taking R or r as a reference length.

If the frequency for the $n = 0$ mode is assumed to be fixed, the co-ordinates of Figs. 4 and 5 directly yield the size of the electrode radius. The smaller the value of ζ, the larger the size and the worse the mode separation. If both the frequency and the size are fixed, Figs. 4 and 5 show the required value of ζ and Fig. 6 gives the difference of the mode separation between two types.

As an illustration, for a fixed frequency of 1000 Mc/s, the ratio R/r and the mode separation are given as a function of the interaction-space radius R_0 in Fig. 7, assuming $C_c = 0$. We see that the mode separation is always better for the inverted magnetron structure. Figure 8 gives the ratio R/r and the mode separation as a function of the geometric mean radius $R_m = \sqrt{(Rr)}$. We see that for a fixed effective size the conventional magnetron has a better mode separation.

The results described in Figs. 4 and 5 approximate the frequencies quite well even on the basis of a rough estimate of C_c. Many experimental results yield accuracies of the order of a few per cent.

4. Equivalent-circuit Representation

4.1. *Equivalent Circuits for a Cylindrical Magnetron Structure with Bars*

We can synthesize an equivalent circuit for the waves travelling along the bars of the structure of Figs. 1 or 2 if we assume that neighbouring bars constitute a parallel-wire transmission line periodically loaded by the reactance of the cavities between the parallel plates. This equivalent circuit allows the circuit characteristics to be determined (if the structure is used as an amplifier) and also the resonant frequencies of the synthesized structure (used as an oscillator). It is shown in Fig. 9. The symbol L_{eq} denotes a reactance that changes from inductive to capacitive as the frequency is changed.[6] The symbol may be replaced by C_{eq}, defined by $\omega^2 L_{eq} C_{eq} = -1$. The symbol C in Fig. 9 denotes the capacitance between bars in an axial length Δ (Figs. 1 and 2) and

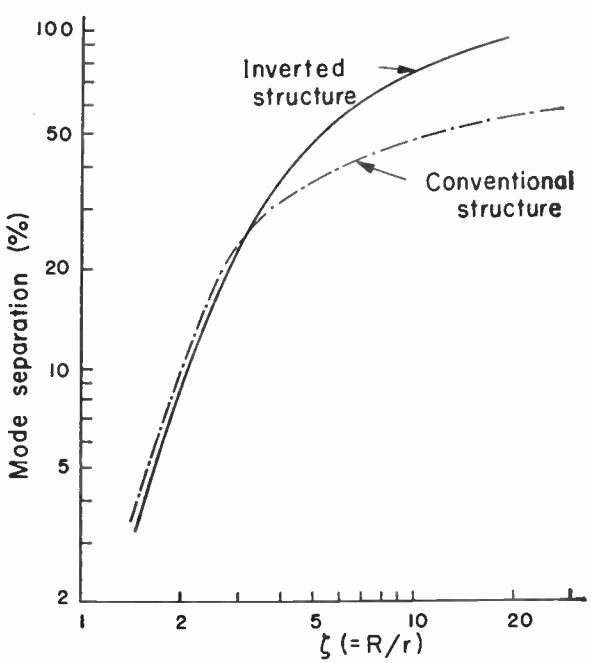


Fig. 6. Degree of mode separation.

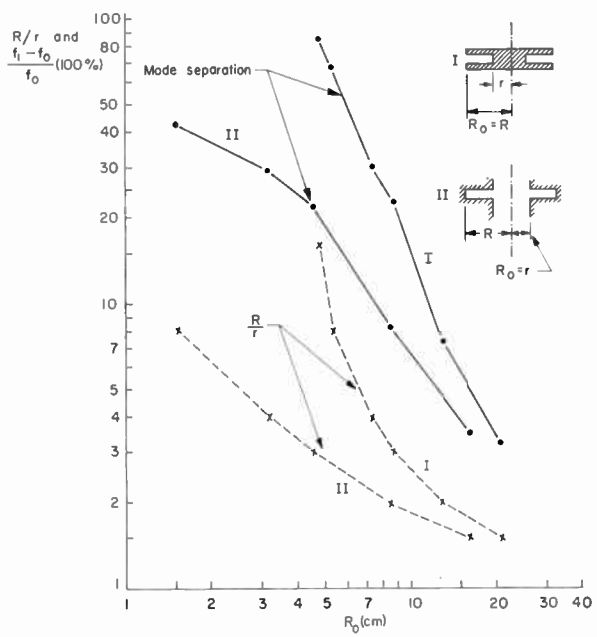


Fig. 7. Mode separation (solid line) and R/r ratio (dashed line) at 1 Gc/s as a function of R_0.

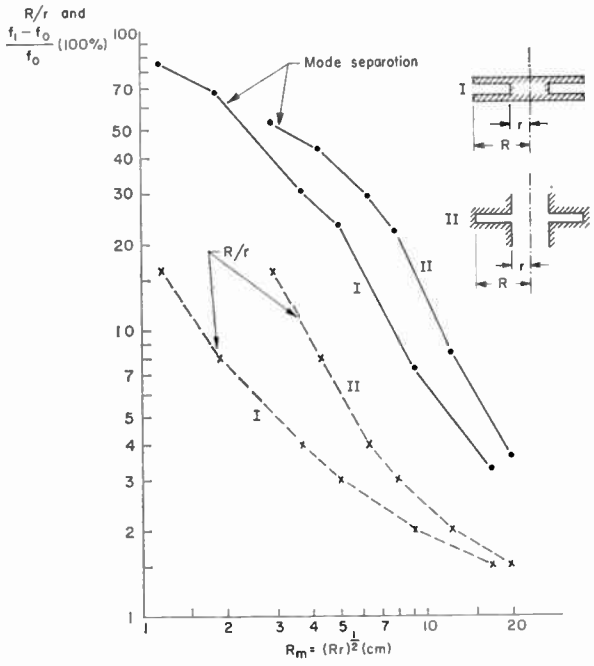


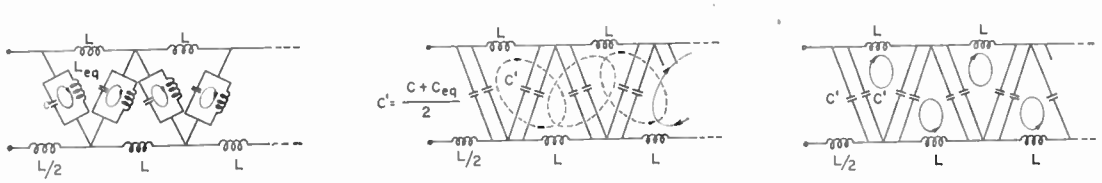
Fig. 8. Mode separation (solid line) and R/r ratio (dashed line) at 1 Gc/s as a function of mean radius $(Rr)^{\frac{1}{2}}$.

should also contain the effect of the plate edges. The symbol L represents the inductance of bars per section (Δ), which can be usually estimated, at least approximately.

Figure 9 shows lumped continuous transmission lines made up of elementary four-terminal networks, from which we can estimate the behaviour of the circuit. Such an analysis may be taken as a good approximation for $n = 0$ mode. However, it is not as good an approximation for higher-order modes because of the variation of field along the circumference of the plates.

Under these considerations the phase diagram and the characteristic impedance as a function of frequency have been calculated for an inverted magnetron structure whose dimensions are illustrated in Fig. 10, where the results of this calculation are shown both for the $n = 0$ (TM_{010} and TM_{020}) and the $n = 1$ (TM_{011}) modes. The frequency f is one co-ordinate and the phase parameter per unit section $\beta\Delta$ is the other. Also plotted along the same axis as $\beta\Delta$ is the characteristic impedance Z_c divided by the bar number N (here, 60).

Points A_0 and A_1 denote the conditions for no energy transmission along the bars (parallel resonance: phase velocity $v_p = \infty$, group velocity $v_g = 0$); and C and L_{eq} have the value corresponding to parallel resonance for each mode (Fig. 9(a)). Points B_0 and B_1 denote the conditions for series resonance (Fig. 9(b)). Two groups of circuit elements are possible for such a series circuit, as shown in Fig. 9(b) by the solid and dotted lines. Points C_0 and C_1 denote another resonant condition that can occur for each mode (Fig. 9(c)). The phase difference per section is π in Fig. 9(b) and 2π or zero in Fig. 9(c).



(a) Parallel resonance. (b) Series resonance. (c) Combined resonance.

Fig. 9. Equivalent-circuit representation for various resonant conditions.

Measurements made with a synthesized structure similar to the model of Fig. 10 show fairly good agreement between the lower frequencies calculated from the figure and those measured. It is found from Fig. 10 that the mode separation of the synthesized structure may be simply estimated from that of an individual cavity, even if the number of cavities amounts to around ten. The ω-β diagram shows that over a limited frequency range the dispersion is very small.

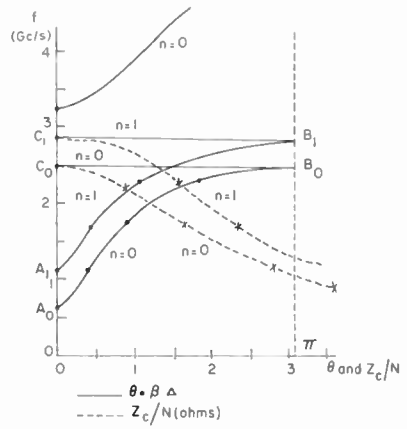


Fig. 10. Phase characteristics and total characteristic impedance for structure of Fig. 1.

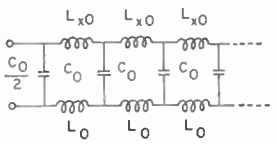
4.2. *Equivalent Circuits for Linear Magnetron Structures*

4.2.1. For the $n = 0$ mode

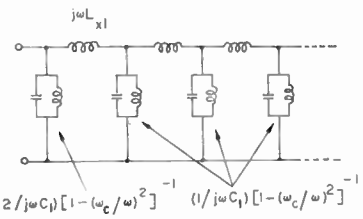
The equivalent circuit for the structure of Fig. 3 is shown in Fig. 11(a). The expression for the loading reactances L_{eq0}'s of the unit cell is as easily derived as in the previous case.[6] The capacitance C_0 between the cathode cylinder and the anode per unit section, and the inductance L_0 due to the interaction space per unit section, can be likewise simply derived, yielding the phase characteristics and the characteristic admittance Y_c for waves propagating in the axial direction, from the lumped-circuit equivalent.

4.2.2. For the $n = 1$ mode

A good equivalent-circuit representation is more difficult in this case because of the field distribution. An equivalent circuit that is a good approximation to the structure is shown in Fig. 11(b). The quantity ω_c is the lowest resonant frequency of the TE coaxial-line mode. The effective capacitance C_1 is determined from the maximum value of electric energy per unit volume stored in the space between the anode and the cathode for the $n = 1$ mode. The circuit of Fig. 11(b) has its cut-off at ω_c. The effective loading inductance $L_{eq\,1}$ can also be determined from energy considerations.



(a) For $n = 0$ mode.



(b) For $n = 1$ mode.

Fig. 11. Equivalent circuits.

4.2.3. Numerical calculation and experiment

The phase characteristics and the characteristic admittances for the $n = 0$ and $n = 1$ modes for a model representing the structure of Fig. 3 are shown in Fig. 12. Experiments show that all the lower-mode resonances of this structure can be predicted from the phase characteristic diagram with an error smaller than a few per cent, attesting to the accuracy of the calculation.[6] The ω-β diagram is characterized by a relatively abrupt initial rise, followed by a slower changing slope.

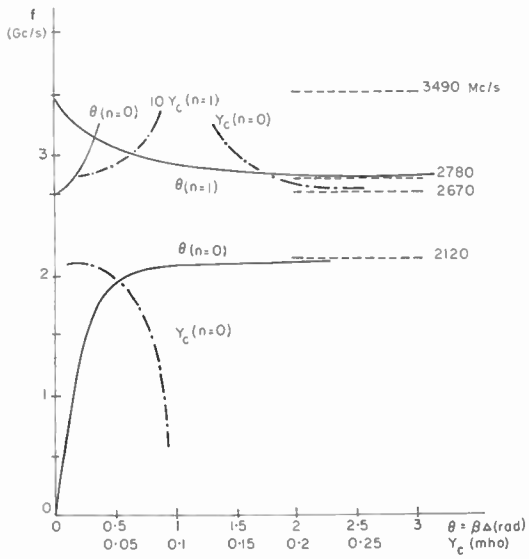


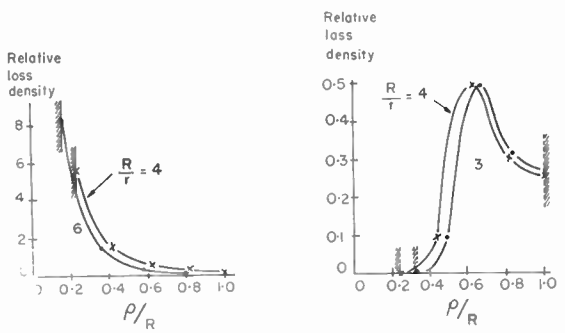
Fig. 12. Phase characteristics and characteristic impedance for structure of Fig. 3. $R/r = 4$, $r = 1{\cdot}5$ cm, $a = 1{\cdot}2$, $\delta = 0{\cdot}1$ cm, $\Delta = 0{\cdot}6$ cm.

5. Additional Considerations

5.1. *Cavity Losses and Q*

Distribution of loss along the radial plates of an individual cavity and the approximate Q of the structure are important, since the gap length of a cavity might have to be less than a few millimetres for a reasonable frequency range owing to multipactor considerations; concentration of loss at a particular location must be avoided for high-power applications.

The loss-density distributions on the plate surfaces have been calculated[6] for certain models from the electromagnetic field in the cavity (Fig. 13). The loss density for an inverted magnetron structure rises to a relatively high value near the central cylinder, which could lead to difficulties in cooling a high-power magnetron structure. By way of comparison, the losses for conventional magnetron structures are relatively spread out over the plates; however, one can visualize circumstances under which heat concentrated in the central structure might be more easily dealt with than distributed heat.



(a) For inverted structures. (b) For conventional structures.

Fig. 13. Loss-density distribution over plate surfaces. (ρ is the radial co-ordinate.)

Values of Q of the order of 1000 may be expected for a gap length of 1·5 mm and a radial size of several centimetres, and are roughly proportional to the gap length and inversely proportional to the root of the radial size for a given R/r ratio. The total interaction impedance as seen from the open end of the cavity is of the order of a few thousand ohms for the same structures and is approximately proportional to the square of the gap length and inversely proportional to the 3/2 power of the radial size.

5.2. *Improvement of Mode Separation*

Mode separation can be improved by changing the gap length as a function of radius, as shown in Fig. 14; the frequency dependence changes from mode to mode and is different from the uniform-gap case.

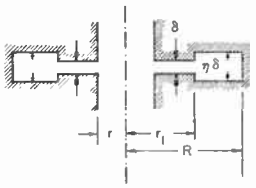


Fig. 14. Structure used for mode-separation improvement.

The exact field solution for this case is rather involved because of the additional boundary conditions at r_1, where the gap length changes. With the assumption that current and potential difference are continuous at r_1, calculations show that for $\eta > 1$ (Fig. 14) the mode separation may be strongly increased at the expense of frequency. Conversely, a structure with $\eta < 1$ may yield a higher frequency for a given size.

5.3. *Other Considerations*

An important design consideration is the prevention of multipactor action. Since the r.f. voltage between the two radial plates constituting an individual cavity changes along the radial path from the maximum at the open rim to zero at the closed rim, one procedure might be to choose the value of δ/λ_0 so large that the maximum amplitude of the r.f. voltage may be smaller than the multipactor voltage for an appropriate high-order mode. Another (and more effective) procedure is to choose δ/λ_0 small enough to allow the multipactor relation only under the condition of trivial r.f. voltage, for which no current amplification occurs.

The multipactor phenomenon might well limit the design of the structures seriously, particularly at the higher frequencies and/or higher voltages. Structures such as that of Fig. 14 could be expected to be easily effective in that regard. For instance, the gap length corresponding to the closed-edge side of a cavity may be selected as large as necessary, and that corresponding to the open-edge side may be chosen as small as necessary.

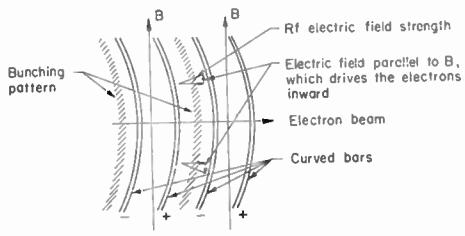


Fig. 15. Developed view of modified structure.

Table 1

Type of anode structure		Wavelength (cm)	Mode separation (%)	Anode-surface radius at interaction space (cm)	Anode-surface radius at other side (cm)	Remarks
HP10V (double ring strap)		10·7	5	1·5	2·86	From Collins[7]
BM50 (single ring strap)		3·2	27·5	0·128	0·58	
New anode structures	Inverted type	10·7	5	4·1	2·41	$C_c\delta/\varepsilon_0 R = 0{\cdot}2$
		3·2	27·5	0·66	0·194	
		3·2	5	1·14	0·67	
	Conventional type	10·7	5	3·4	5·61	$C_c\delta/\varepsilon_0 r = 0{\cdot}2$
		3·2	27·5	0·35	1·26	
		3·2	5	1·1	1·82	

A slight structural modification may have additional advantages. Suppose that every bar is curved slightly in the direction of electron-beam rotation (Fig. 15). The component of the r.f. electric field parallel to the d.c. magnetic field might then be expected to provide a confining effect on the electron bunches, thus improving the bunching action.

Finally, in any complete cold-cathode-magnetron design, an effective way must be developed of damping out the extraneous modes that may be expected to occur in the interaction space because of the cold-cathode configuration. The damping might be accomplished by the use of an axially-slotted cathode, so that a circumferential current would encounter a strong impedance.[2]

6. Conclusions

Resonant frequencies and mode separation have been determined for an inverted magnetron structure and for a conventional one. These structures were chosen because of the desirability of having appropriate circuits for high-power magnetrons from the viewpoint of having the largest possible interaction volume. The proposed structures can be larger than conventional structures such as the 'strapped' or the 'rising sun' for a fixed frequency (Table 1). The structures proposed may be expected to have a high interaction impedance and are easier to fabricate than conventional circuits. Moreover, they may be of interest in connection with travelling-wave amplifiers, in which the waves travel along the axes of symmetry.

7. Acknowledgments

The authors gratefully acknowledge the suggestions made by Profs. D. H. Sloan and A. W. Trivelpiece in the course of the investigation, as well as the technical assistance of O. B. Westwick and J. Tombaugh of the Electronics Research Laboratory, University of California, Berkeley. The work was supported by U.S. Air Force Contract No. AF19(628)-324 and U.S. Army, Navy and Air Force Contract No. AF-AFOSR-139-63.

8. References

1. E. D. Hoag, "Glow Cathodes in Pulsed Magnetic Fields". Electronics Res. Lab., University of California, Berkeley, Calif., Sci. Rept. No. 4, Issue 239, Series 60; 15th June 1959,
2. C. W. Hartman, "Production and Interactions of Electron Beams in Crossed Fields", Electronics Res. Lab., University of California, Berkeley, Calif., Sci. Rept. No. 10, Issue 325, Series 60; 31st October 1960.
3. M. Chamran, "Electron Beam in the Cold-Cathode Magnetron", Electronics Res. Lab., University of California, Berkeley, Calif., Sci. Rept. No. 18, Issue 453, Series 60; 15th June 1962.
4. S. L. Mercer, "Cold-Cathode Power Tube", Electronics Res. Lab., University of California, Berkeley, Calif., Rept. No. 64–24; 23rd July 1964.
5. Y. Ikeda, "Behavior of the Space Charge in a Plasma Magnetron", Electronics Res. Lab., University of California, Berkeley, Calif., Sci. Rept. No. 17, Issue 433, Series 60; 15th June 1962.
6. Y. Ikeda, "Anode Structures for Cold-Cathode High-Power Magnetron", Electronics Res. Lab., University of California, Berkeley, Calif., Sci. Rept. No. 20, Issue 455, Series 60; 30th June 1962.
7. G. B. Collins (ed.), "Microwave Magnetrons", pp. 739–96. (McGraw-Hill, New York, 1948.)

Manuscript first received by the Institution on 31st July 1964, and in final form on 6th November 1964. (Paper No. 599.)

U.K.–Belgium Submarine Cable

Provision of Submerged Transistor Repeaters

Submerged transistor repeaters have recently been inserted in a submarine cable between St. Margaret's Bay in East Kent and La Panne in Belgium to increase its capacity from 216 to 420 telephone circuits. These are the first submerged repeaters using transistors to be used in an international submarine telephone cable and the first time that so many circuits have been operated on such a cable.

The cable was laid in 1948 and is of unusual design in that the polythene dielectric incorporates an air space obtained by a helix of polythene cord laid over the centre conductor before extruding the main polythene dielectric. The coaxial pair has an internal diameter of 1·7 in for its outer conductor. The route length is about 47·5 n.m.

The 216 telephone circuits obtained on the cable before the repeaters were inserted were transmitted from the U.K. to Belgium in the band 60–956 kc/s and in the band 1152–2048 kc/s in the reverse direction. At La Panne the 216 circuits were connected at line frequency range (60–956 kc/s) to two coaxial tubes of a conventional coaxial land system to Brussels. The new 420 circuit system uses line frequencies in the bands 312–2044 kc/s and 2544–4276 kc/s and, as for the earlier system, the circuits are extended to Brussels in the line frequency range 312–2044 kc/s over the coaxial land system. The highest working frequency over the submarine cable is about 4·3 Mc/s, at which frequency the total cable loss is about 130 dB. The two submerged repeaters amplify the two directions of transmission using common amplifiers and have a gain of about 45 dB at 4·3 Mc/s. They are inserted in the cable at approximately 16 n.m. and 32 n.m. from the La Panne terminal.

The terminal transmission and power feeding equipment are fully transistorized. Two pilot frequencies located near the edges of the traffic band in each direction of transmission continuously monitor the system performance. Variable equalizers are provided at the terminal stations to maintain the performance with changes of cable attenuation due to seasonal variations of sea temperature.

The submerged repeaters and the associated terminal equipment have been designed and manufactured by Submarine Cables Ltd. in association with the Telecommunications Division of Associated Electrical Industries Ltd., who also installed the equipment at the terminal stations. Submarine Cables Ltd. had the overall responsibility for the satisfactory performance of the completed project.

Submerged repeaters using thermionic valves as the active elements in their amplifiers are now in widespread use and their reliability is firmly established. An essential feature of the submerged repeater is the emphasis placed on its reliability, and as one of the factors to achieve this, only components of proven reliability are incorporated. The transistor offers obvious advantages for a submerged repeater because its power requirement and operating voltage are substantially less than those required by a thermionic valve to give comparable performance; these are significant advantages when more circuits involving higher frequencies and closer repeater spacings are required. Much work is being done in the U.K. in this field, including the development of long life transistors suited to the future needs of wide band submerged repeaters on routes much longer than the U.K.–Belgium cable. The repeaters laid in this cable use transistors selected from a batch of high-grade commercial transistors which were the subject of an extensive programme of reliability assessment tests to confirm their suitability for this application.

The repeaters are housed in conventional double-ended cylindrical steel housings which are provided with 0·935 in diameter, solid polythene dielectric, coaxial, armoured cable tails; this latter arrangement enabled well-tried repeater cable gland techniques to be used. Specially developed taper joints between the 0·935 in tail cables and 1·7 in stock cable were made on board the cable ship during the loading in preparation for the repeater laying.

The repeaters were laid by the Post Office cable ship H.M.T.S. *Iris* between 30th October and 2nd November 1964. Fortunately the weather was good and the operation went smoothly; the work involved grappling and lifting the somewhat heavy cable—19 tons per n.m.—and there was a risk that the cable would be heavily sanded, particularly at the La Panne end of the route.

Extensive tests made following the insertion of the repeaters confirm that the system fully meets the specification and provides 420 telephone circuits which adequately meet C.C.I.T.T. recommendations for international telephone circuits.

The completion of this new system is a major step in the development of submerged repeater techniques. Systems are now on order which will use transistor repeaters to provide 480 circuits on several new cables to be laid in the North Sea in 1966. The success of these first transistor repeaters in an international submarine cable providing more circuits than ever before amply demonstrates the lead which British manufacturers have in this field of submarine cable telephony.

U.D.C. 621.396.677

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

By

Professor H. S. HEAPS, B.Sc., M.A.†

AND

C. WADDEN, M.Eng.‡

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

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

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

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

List of Symbols

$A+jB$	normalized frequency response of processing networks
$K(\theta)$	(signal/noise)^2 for single receiver
$p_s(X, t)$	signal incident at position X at time t
$p_n(X, t)$	noise at X at time t
$p_n(\theta, \phi, t)$	plane wave from direction θ, ϕ
$P_{s\tau}^2$	mean square output signal averaged over duration τ
P_s^2	mean square output signal
P_n^2	mean square output noise
$r(X, t)$	unit impulse response of the network connected to the receiver at X
$R(X, \omega)$	frequency response to radiation incident at X
T	half averaging time
T_{12}	time for sound to travel from X_2 to X_1
$V_n(t)$	output noise
$V_s(t)$	output signal
$V_s(\omega)$	frequency distribution of output signal $V_s(t)$
x_n, y_n, Y, Z	$\omega X/c$
$\lambda_d, \lambda_m, \lambda_s$	maximum ratio of (signal/noise)^2
θ, ϕ	direction of plane wave
θ_c, ϕ_c	direction of signal incident from a single direction
$\phi(X_1, X_2, t)$	space correlation function
$\Phi(X_1, X_2, \omega$	Fourier transform of $\phi(X_1, X_2, t)$
$\Psi_n(\theta, \phi, t)$	time correlation function of plane waves of noise

1. Introduction

In the design of hydrophone or antenna arrays there arises the problem of how to process the received signals so as to obtain an optimum form of output. Possible criteria for judgement of optimum are minimum beam width, minimum level of side-lobes, and maximum signal/noise ratio. The latter criterion is the one adopted in the present paper.

An array in which the outputs from the various elements are added and then time averaged may be termed an additive array. Much of the early theory relating to antenna design dealt with the design of additive arrays to produce given directivity patterns. W. W. Hansen and J. R. Woodward[1] described the design of an improved end-fire array in which the exciting currents were kept constant and in which phase delays were introduced into the receiver outputs in order to maximize the total gain.

The directivity pattern of an antenna array may be represented by a complex polynomial. The properties of this polynomial were investigated by S. A. Schelkunoff[2] who noted that the directivity pattern of the array is governed by the location of the zeros of the complex polynomial. By equally spacing the zeros of the polynomial on a unit circle he was able to produce

† Department of Engineering Mathematics, Nova Scotia Technical College; Mathematical Consultant to Naval Research Establishment, Dartmouth, Nova Scotia.

‡ Formerly at Nova Scotia Technical College; now with Computing Devices of Canada Limited.

a design which possessed a reduced major beam width and a reduced minor lobe level.

In general, the directivity pattern of an array consists of a major lobe and several minor lobes of unequal amplitude. C. L. Dolph[3] described a method of design to produce a minimum width of major beam for a given minor lobe level. He concluded that the smallest major lobe width could be obtained by making all of the minor lobes equal to the largest of the minor lobes. Since Chebychev polynomials have a similar property they may be used to calculate the exciting coefficients for a broadside array whose element spacing is greater than one-half of the wavelength of the incident radiation.

H. J. Riblet[4] extended Dolph's procedure to consider a broadside array of an odd number of isotropic elements with spacing less than half the wave length of the incident radiation. Such arrays may be designated as superdirective. G. Sinclair and F. V. Cairns[5] developed a procedure for the optimum design for superdirective broadside arrays with even numbers of elements.

Because of the progressive phase delay inherent in an end-fire array the Dolph-Chebychev pattern is not optimum. R. H. Duhamel[6] produced an optimum design for an end-fire array of odd numbers of elements with spacing less than half the incident wavelength. The procedure used was an extension of Riblet's work and produces designs that are superior to those produced by the method of either Hansen, Woodward, or Schelkunoff.

The Dolph-Chebychev procedure becomes very laborious for arrays with large numbers of elements. To overcome this difficulty D'Barbiere[7] manipulated the equations for the Dolph-Chebychev pattern into a form that lends itself to rapid calculation. R. J. Stegen[8] further improved the equations by equating the directivity pattern to a Fourier series whose coefficients may be easily calculated.

By adding progressive phase delay to an optimum broadside pattern D. R. Rhodes[9] was able to produce an end-fire array with one major lobe. This lobe, however, was considerably wider than that of an end-fire array with two lobes.

R. L. Pritchard[10,11,12] applied the previous work of Dolph and Riblet to the design of hydrophone arrays. The procedure was extended to steered arrays and it was shown that in general optimum Chebychev patterns cannot be steered effectively by the addition of phase delay. D. G. Tucker[13] noted that the directivity pattern is similar to a $(\sin x)/x$ function, and he suggested a design procedure based upon superposition of $(\sin x)/x$ functions spaced along the x axis. In practice this can be obtained by using tapped delay lines.

A multiplicative, or time-averaged product (t.a.p.) array, is an array where the outputs from the elements or sections of elements, are multiplied and then time averaged to give the output of the array. A. Berman and C. S. Clay[14] showed that the directivity pattern obtained with a linear additive array could be obtained by use of a multiplicative array with a smaller number of elements. D. C. Fakley[15] analysed multiplicative arrays and showed that, from the point of view of signal to noise or target discrimination, a t.a.p. array has no advantage over an intraclass correlator.

It has been shown that multiplicative arrays have a higher directivity than linear additive arrays. V. G. Welsby and D. G. Tucker[16] analysed multiplicative arrays from the criteria of signal/noise performance, directivity, and target discrimination. These criteria are related by the directivity pattern for linear additive arrays but may not be for multiplicative arrays. The signal/noise ratio of a multiplicative array is less than that of a linear additive array but higher than that of a superdirective array with the same directivity.

The use of multiplicative arrays for angular resolution of targets in sonar systems was investigated by V. C. Welsby.[17] The results showed that multiplicative arrays have a greater bearing accuracy and are more capable of identifying closely spaced targets than are linear additive arrays.

D. G. Tucker[18] discussed multiplicative receiving arrays as used in sonar and radio astronomy. A two-part multiplicative array was shown to have approximately twice the bearing accuracy of a linear additive array with the same number of elements. Crossed arrays which have a common central element may have their secondary lobes almost completely eliminated.

An important factor in the design of any receiving system is the effective signal/noise ratio, that is the ability of the system to detect small signals in the presence of noise. The signal/noise gain of a hydrophone array is defined by V. G. Welsby[11] to be the signal/noise ratio of the array divided by the signal/noise ratio of a single receiver. For a line array the gain is proportional to the effective length of array. The signal/noise performance of a superdirective array is much less than that of an ordinary array. D. G. Tucker[20] showed that, although the directivity of a superdirective array is better than that of an ordinary one, the signal/noise performance is much worse and causes an overall degradation in performance.

The signal/noise performance of various arrays was investigated by D. G. Tucker.[21] He defined as noise factor of an array the signal/noise ratio of a linear plane array divided by the signal/noise ratio of the array being considered. The noise factor was shown

to be related to the directional pattern by the relation

$$\text{N.F.} = \frac{L^{\frac{1}{2}}\left[\int_{-\infty}^{\infty} |D(x)|^2 \, dx\right]^{\frac{1}{2}}}{\int_{-\infty}^{\infty} T(r)\, dr}$$

where

L is the length of the array,

$D(x)$ is the directional pattern,

$T(r)$ is the tapering function for the array.

In a previous paper H. S. Heaps[22] described a method for determination of the optimum spacing of hydrophone elements, and the optimum linear combination of their outputs, in order to give a maximum peak signal output in comparison to the mean noise background. An integral equation relating the space correlation function of the noise, the impulse response of the array, and the signal pressure, was derived. The solution of this equation gives the optimum response for each element.

Although many designs have been produced for optimizing the directivity patterns of arrays, and studies have been conducted on the signal/noise characteristic of various arrays, little work has been done on the design of hydrophone arrays for optimum signal/noise ratio. The purpose of the present paper is to deduce the theory of the optimization of hydrophone arrays when the optimization is with respect to signal/noise ratio. Several hydrophone arrays are designed and compared from the point of view of signal/noise ratio as a function of signal direction.

2. Mathematical Representation of Array

It is convenient to consider a general configuration of small receivers each occupying a space dX in the neighbourhood of a point X. The total region occupied by the receivers is

$$\int dX \qquad \text{.....(1)}$$

The region dX may be of one, two, or three dimensions, and may be continuous or discrete. For example, for the treatment of an ideal plane transducer the region $\int dX$ may be chosen as the aperture plane and the integration (1) is two dimensional. For treatment of a line receiver the region $\int dX$ may be chosen as the line of the receiver and the integral (1) is one dimensional. To treat a discrete array of receiving elements that occupy small volumes $dX_1, dX_2, \ldots, dX_n$ the integration (1) is to be replaced by a summation over dX_1 to dX_n.

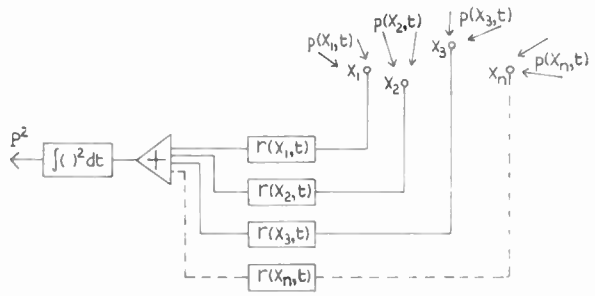


Fig. 1. System of receivers and processing networks in a linear array.

Suppose the receiving element of volume dX at X receives an acoustic, or electromagnetic, radiation of noise of amplitude $p_n(X, t)$. The output of the receiver is fed to a linear processing network whose output at time t is

$$dX \int p_n(X, t-t_1) r(X, t_1)\, dt_1 \qquad \text{.....(2)}$$

where $r(X, t)\, dX$ denotes the output of the network in the special case that $p_n(X, t)$ is a unit impulse at time $t = 0$. The integration with respect to time is from $-\infty$ to ∞ although for a physically realizable process $r(X, t)$ must be zero for negative values of t. The impulse response $r(X, t)$ includes the effect of the response of the receiving element as well as the response of the processing network.

If the outputs of all the processing networks are added then the resultant output noise is

$$V_n(t) = \int\int p_n(X, t-t_1) r(X, t_1)\, dX\, dt_1 \qquad \text{.....(3)}$$

The mean square output noise is

$$\begin{aligned} P_n^2 &= \lim_{T\to\infty} (1/2T) \int_{-T}^{T} |V_n(t)|^2 \, dt \\ &= \int\int\int\int r(X_1, t_1) r^*(X_2, t_2) \times \\ &\qquad \times \phi_n(X_1, X_2, t_2 - t_1)\, dX_1\, dX_2\, dt_1\, dt_2 \qquad \text{.....(4)} \end{aligned}$$

in which $\phi_n(X_1, X_2, t_2 - t_1)$ denotes the function

$$\begin{aligned} \phi_n(X_1, X_2, t_2 - t_1) &= \lim_{T\to\infty} (1/2T) \int_{-T}^{T} p_n(X_1, t-t_1) \times \\ &\qquad \times p_n^*(X_2, t-t_2)\, dt \\ &= \lim_{T\to\infty} (1/2T) \int_{-T}^{T} p_n(X_1, t) \times \\ &\qquad \times p_n^*(X_2, t+t_1-t_2)\, dt \qquad \text{.....(5)} \end{aligned}$$

The superscript asterisk denotes complex conjugate.

For a discrete set of point receivers the system of receivers and processing networks is as shown in Fig. 1 for an incident pressure p.

The function $\phi_n(X_1, X_2, t)$ is the cross-correlation function of the incident radiation at positions X_1 and X_2. Its Fourier transform is the function

$$\Phi_n(X_1, X_2, \omega) = \int \phi_n(X_1, X_2, t) \exp(-j\omega t)\, dt \qquad \text{.....(6)}$$

Use of the convolution theorem enables eqn. (4) to be written as

$$\begin{aligned} P_n^2 &= (1/2\pi) \int\int\int R(X_1, \omega) R^*(X_2, \omega) \times \\ &\qquad \times \Phi_n(X_1, X_2, \omega)\, dX_1\, dX_2\, d\omega \qquad \text{.....(7)} \end{aligned}$$

where

$$R(X,\omega) = \int r(X,t)\exp(-j\omega t)\,dt \qquad \ldots\ldots(8)$$

is the frequency response of the hydrophone and processing network that responds to radiation incident over the volume dX at the position X. $R^*(X,\omega)$ denotes the complex conjugate of $R(X,\omega)$.

The noise field may be termed uniform if $\phi_n(X_1, X_2, t)$ is a function only of t and the distance between the points X_1 and X_2. For a uniform noise field $\phi_n(X_1, X_1, t) = \phi_n(X_2, X_2, t)$.

3. Optimum Array with Continuous Signal

The mean square output signal of the system shown in Fig. 1 is

$$P_s^2 = \lim_{T\to\infty}(1/2T)\int_{-T}^{T}|V_s(t)|^2\,dt$$
$$= (1/2\pi)\int\int\int R(X_1,\omega)R^*(X_2,\omega)\times$$
$$\times\Phi_s(X_1,X_2,\omega)\,dX_1\,dX_2\,d\omega \qquad \ldots\ldots(9)$$

For ease of detection of the signal the various processing networks should be chosen to maximize P_s^2/P_n^2. Accordingly let $R(X,\omega)$ be chosen to maximize P_s^2/P_n^2 and let the maximum value be denoted by λ_m. It is shown in the Appendix that the equation

$$\int[\Phi_s(X_1,X_2,\omega)-\lambda_m\Phi_n(X_1,X_2,\omega)]\times$$
$$\times R^*(X_2,\omega)\,dX_2 = 0 \qquad \ldots\ldots(10)$$

must then be satisfied for each value of X_1 and ω. The maximum value of P_s^2/P_n^2 is equal to the maximum value of λ_m for which a non-zero function $R(X,\omega)$ may be found to satisfy eqn. (10) for all values of X_1 and ω.

In general the solution λ_m of eqn. (10) is a function of ω. The maximum value of λ_m occurs at a particular value, or set of values, of ω. The maximum value of P_s^2/P_n^2 results if detection is at these particular values of ω, the function $R(X,\omega)$ being chosen to equal zero at other values. Alternatively, if for each ω the function $R(X,\omega)$ is chosen to maximize λ_m in (10), then there results a system which enhances in a maximum manner the signal/noise ratio at each frequency.

3.1. *The Two-element Array*

In the special instance when the array consists of two hydrophones situated at positions X_1 and X_2, then eqn. (10) signifies that the determinant

$$\begin{vmatrix}\Phi_s(X_1,X_1,\omega)-\lambda_m\Phi_n(X_1,X_1,\omega) & \\ \Phi_s(X_1,X_2,\omega)-\lambda_m\Phi_n(X_1,X_2,\omega) & \end{vmatrix}$$

It is then possible to determine the ratio

$$R(X_2,\omega)/R(X_1,\omega_1).$$

If both the noise and signal fields are uniform it follows from eqn. (11) that

$$\frac{\lambda_m}{\Phi_s(X,X,\omega)/\Phi_n(X,X,\omega)} = \frac{1-sn\pm(s-n)}{1-n^2} \qquad \ldots\ldots(12)$$

where s and n denote $\Phi_s(X_1, X_2, \omega)/\Phi_s(X, X, \omega)$ and $\Phi_n(X_1, X_2, \omega)/\Phi_n(X, X, \omega)$ respectively. Thus the maximum value of λ_m is given by

$$\frac{\lambda_m}{\Phi_s(X,X,\omega)/\Phi_n(X,X,\omega)} = \begin{cases}(1-s)/(1-n) \text{ if } s<n\\ (1+s)/(1+n) \text{ if } s>n\end{cases}$$
$$\ldots\ldots(13)$$

The quantity $\Phi_s(X, X, \omega)/\Phi_n(X, X, \omega)$ is the signal/noise ratio P_s^2/P_n^2 which results from a single receiving element. Thus the right hand side of eqn. (13) is the factor by which the use of two hydrophones may increase the signal/noise ratio. Unless s and n are equal to zero the right-hand side of eqn. (13) exceeds unity and so the signal/noise ratio may be increased by the use of two receivers instead of a single one.

4. Optimum Array with Pulsed Signal

If the signal has the form of a pulse which contains a finite amount of energy within a finite time interval then the quantity P_s^2 of eqn. (9) is zero. The mean square output signal may however be computed as an average over a finite duration τ, thus leading to consideration of the quantity

$$P_{s\tau}^2 = (1/\tau)\int_t^{t+\tau}|V_s(t)|^2\,dt$$
$$= (1/\tau)(1/4\pi^2)\int_t^{t+\tau}dt\int\int\int\int P(X_1,\omega_1)P^*(X_2,\omega_2)\times$$
$$\times R(X_1,\omega_1)R^*(X_2,\omega_2)\times$$
$$\times\exp j(\omega_1-\omega_2)t\,dX_1\,dX_2\,d\omega_1\,d\omega_2$$
$$= (1/4\pi^2)\int\int\int\int P(X_1,\omega_1)P^*(X_2,\omega_2)R(X_1,\omega_1)\times$$
$$\times R^*(X_2,\omega_2)G(\omega_1-\omega_2)\,dX_1\,dX_2\,d\omega_1\,d\omega_2$$
$$\ldots\ldots(14)$$

where

$$G(\omega) = \frac{[\exp(j\omega t)-1]}{j\omega t}\exp(j\omega t) \qquad \ldots\ldots(15)$$

and

$$P(X,\omega) = \int p_s(X,t)\exp(-j\omega t)\,dt \qquad \ldots\ldots(16)$$

The signal field may be termed uniform if $|P(X_1, \omega)| = |P(X_2, \omega)|$ for all choices of X_1 and X_2.

By a method similar to that used to derive eqn. (10) it may be shown that if $R(X, \omega)$ is chosen to maximize $P_{s\tau}^2/P_n^2$, and if the maximum value is denoted by λ_d,

$$\begin{vmatrix}\Phi_s(X_2,X_1,\omega)-\lambda_m\Phi_n(X_2,X_1,\omega) & \\ \Phi_s(X_2,X_2,\omega)-\lambda_m\Phi_n(X_2,X_2,\omega) & \end{vmatrix} = 0 \qquad \ldots\ldots(11)$$

then for each X, ω

$$P(X,\omega)\int\int P^*(X_1,\omega_1)G(\omega-\omega_1)R^*(X_1,\omega_1)\,dX_1\,d\omega_1$$
$$-2\pi\lambda_d\int R^*(X_1,\omega)\Phi_n(X,X_1,\omega)\,dX_1$$
$$= 0 \qquad \ldots\ldots(17)$$

4.1. *Monochromatic Signal*

It may be noted that if the signal energy is concentrated in a single frequency ω then eqn. (17) becomes

$$\int [P(X,\omega)P^*(X_1,\omega)-2\pi\lambda_d\Phi_n(X,X_1,\omega)]\times \times R^*(X_1,\omega)\,dX_1 = 0 \quad \text{.....(18)}$$

Since $P(X,\omega)\,P^*(X_1,\omega)$ is the Fourier transform of the time correlation function of $p_s(X,t)$ and $p_s(X_1,t)$ the solution of eqn. (18) is equivalent to solution of eqn. (10) at the single frequency ω. An alternative form of eqn. (17) results from the relation

$$V_s(\omega) = \int P(X_1,\omega)R(X_1,\omega)\,dX_1$$

where $V_s(\omega)$ is the frequency distribution of the signal output $V_s(t)$. Then eqn. (18) becomes

$$P(X,\omega)V_s^*(\omega) = 2\pi\lambda_d\int \Phi_n(X,X_1,\omega)R^*(X_1,\omega)\,dX_1$$

The maximum signal/noise ratio is then

$$\frac{P_{st}^2}{P_n^2} = \frac{(1/2\pi)\left|\int P(X,\omega)Y(X)\,dX\right|^2}{\int\int \Phi_n(X_1,X_2,\omega)Y(X_1)Y^*(X_2)\,dX_1\,dX_2}$$

where $Y(X)$ denotes $\lambda_d R(X,\omega)/V_s^*(\omega)$ and may be determined as the solution of the equation

$$2\pi\int \Phi_n(X,X_1,\omega)\,Y^*(X_1)dX_1 = P(X,\omega).$$

In the special instance that the array consists of two hydrophones at positions X_1 and X_2, eqn. (18) reduces to

$$\begin{vmatrix} |P(X_1,\omega)|^2-2\pi\lambda_d\Phi_n(X_1,X_1,\omega) & P(X_1,\omega)P^*(X_2,\omega)-2\pi\lambda_d\Phi_n(X_1,X_2,\omega) \\ P(X_2,\omega)P^*(X_1,\omega)-2\pi\lambda_d\Phi_n(X_2,X_1,\omega) & |P(X_2,\omega)|^2-2\pi\lambda_d\Phi_n(X_2,X_2,\omega) \end{vmatrix} = 0 \quad \text{.....(19)}$$

If both the noise and signal are uniform then the non-zero solution of (19) is given by

$$\frac{\lambda_d}{|P(X,\omega)|^2/2\pi\Phi_n(X,X,\omega)} = \frac{2(1-n\cos\theta)}{1-n^2} \quad \text{.....(20)}$$

in which n denotes $\Phi_n(X_1,X_2,\omega)/\Phi_n(X,X,\omega)$ and θ denotes the phase difference between $P(X_1,\omega)$ and $P(X_2,\omega)$.

4.2. *Observation at a Single Time*

If the signal is not restricted to be of a single frequency, but the output is observed at a single time t, then $\tau = 0$ and $G(\omega) = \exp(j\omega t)$. Equation (17) takes the form

$$P(X,\omega)\exp(j\omega t)\int\int P^*(X_1,\omega_1)R^*(X_1,\omega_1)\times \times\exp(-j\omega_1 t)\,dX_1\,d\omega_1-2\pi\lambda_t\int\Phi_n(X,X_1,\omega)\times \times R^*(X_1,\omega)\,dX_1 = 0 \quad \text{.....(21)}$$

and hence

$$P(X,\omega)V_s^*(t)\exp(j\omega t)-\lambda_t\int\Phi_n(X,X_1,\omega)\times \times R^*(X_1,\omega)\,dX_1 = 0 \quad \text{.....(22)}$$

It is convenient to define a quantity $Y(X,\omega)$ by the relation

$$Y(X,\omega) = [\lambda_t/V_s(t)]R(X,\omega) \quad \text{.....(23)}$$

Then $Y(X,\omega)$ may be determined as the solution of the integral equation

$$\int Y^*(X_1,\omega)\Phi_n(X,X_1,\omega)\,dX_1 = P(X,\omega)\exp(j\omega t) \quad \text{.....(24)}$$

The maximum value of P_{st}^2/P_n^2 may be derived from eqns. (7), (14) and (24) as:

$$\frac{V_s(t)^2}{P_n^2} = \frac{1}{2\pi}\frac{\int\int\int\int P(X_1,\omega_1)P^*(X_2,\omega_2)Y(X_1,\omega_1)Y^*(X_2,\omega_2)G(\omega_1-\omega_2)\,dX_1\,dX_2\,d\omega_1\,d\omega_2}{\int\int Y(X_1,\omega)\,P(X_1,\omega)\exp(j\omega t)\,dX_1\,d\omega}$$

$$= \frac{1}{2\pi}\int\int P^*(X_2,\omega_2)Y^*(X_2,\omega_2)\exp(-j\omega_2 t)\,dX_2\,d\omega_2 \quad \text{.....(25)}$$

Solution of eqn. (24) for $Y(X,\omega)$ determines $R(X,\omega)$ apart from a proportionality constant $\lambda_t/V_s(t)$ which is of no significance. Substitution for $Y(X,\omega)$ into eqn. (25) leads gives the maximum possible value of instantaneous peak signal voltage in comparison to the root mean square noise voltage P_n.

5. Analysis of Multiplicative System

A multiplicative system is one in which the output is a time average of the product of the outputs from two separate portions of the array as shown in Fig. 2. Each receiving element is connected to a linear filter as in Section 2.

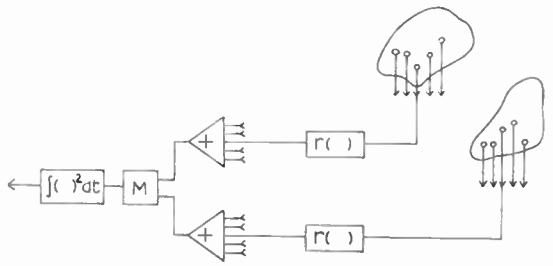


Fig. 2. System of receivers and processing networks in a multiplicative array.

Suppose the two separate portions of the array occupy volumes

$$\int dX \quad \text{and} \quad \int dX' \quad \text{.....(26)}$$

The expression analogous to eqn. (7) for the mean square noise is

$$P_n^2 = (1/2\pi) \int\int\int R(X,\omega)R^*(X',\omega) \times \times \Phi_n(X,X',\omega)\,dX\,dX'\,d\omega \quad .....(27)$$

The expression analogous to eqn. (9) for the mean square signal is

$$P_s^2 = (1/2\pi) \int\int\int R(X,\omega)R^*(X',\omega) \times \times \Phi_s(X,X',\omega)\,dX\,dX'\,d\omega \quad .....(28)$$

Likewise, the expression analogous to eqn. (14) is

$$P_{st}^2 = (1/4\pi^2) \int\int\int\int P(X,\omega_1)P^*(X',\omega_2)R(X,\omega_1) \times \times R^*(X',\omega_2)G(\omega_1-\omega_2)\,dX\,dX'\,d\omega_1\,d\omega_2 \quad .....(29)$$

The values of P_s^2/P_n^2 is a maximum when $R(X,\omega)$ and $R(X',\omega)$ are chosen to satisfy the following two equations which correspond to eqn. (10).

$$\int [\Phi_s(X,X',\omega)-\lambda\Phi_n(X,X',\omega)]R^*(X',\omega)\,dX' = 0 \text{ (for all } X) \quad .....(30a)$$

$$\int [\Phi_s(X',X,\omega)-\lambda\Phi_n(X',X,\omega)]R^*(X,\omega)\,dX = 0 \text{ (for all } X') \quad .....(30b)$$

For a signal in the form of a pulse, and when the output is observed at a single time, the equation analogous to (22) is

$$P(X,\omega)V_s'^*(t)\exp(j\omega t)-\lambda_t \int \Phi_n(X,X',\omega) \times \times R^*(X,X',\omega)\,dX' = 0$$

where

$$V_s'(t) = (1/2\pi) \int P(X',\omega)R(X',\omega)\exp(j\omega t)\,dX'\,d\omega$$

Thus if $Y(X',\omega)$ is defined by the relation

$$Y(X',\omega) = [\lambda_t/V_s'(t)]R(X',\omega)$$

then $Y(X',\omega)$ is the solution of the integral equation

$$\int Y^*(X',\omega)\Phi_n(X,X',\omega)\,dX' = P(X,\omega)\exp(j\omega t) \quad .....(31a)$$

which is analogous to eqn. (24).

The maximum value of P_{st}^2/P_n^2 is then

$$\frac{V_s(t)^2}{P_n^2} = (1/2\pi) \int\int P^*(X',\omega)Y^*(X',\omega) \times \times \exp(-j\omega t)\,dX'\,d\omega \quad .....(31b)$$

which is analogous to eqn. (25).

6. Form of the Space Correlation Function of the Noise

Suppose that in the propagating medium the noise incident upon any unit area consists of a sum of plane waves $p_n(\theta,\phi,t)$ where θ and ϕ are spherical co-ordinates which specify a direction of propagation. At time t let the noise incident upon a receiver at position X_1 be

$$p_n(X_1,t) = (1/4\pi) \int\int p_n(\theta,\phi,t)\sin\theta\,d\theta\,d\phi \quad .....(32)$$

The integrand $p_n(\theta,\phi,t)\sin\theta\,d\theta\,d\phi$ represents the radiation incident from within a solid angle of magnitude $\sin\theta\,d\theta\,d\phi$ centred along the directions θ, ϕ as shown in Fig. 3. The integrations with respect to θ and ϕ are respectively between the limits zero to π and zero to 2π.

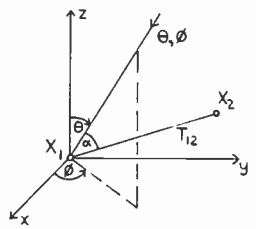


Fig. 3. Geometrical meaning of the angles θ, ϕ, α.

Suppose there is negligible attenuation as each plane wave travels between receivers at X_1 and X_2. The noise incident upon a receiver at X_2 is then

$$p_n(X_2,t) = (1/4\pi) \int\int p_n(\theta,\phi,t+T_{12}\cos\alpha) \times \times \sin\theta\,d\theta\,d\phi \quad .....(33)$$

where $\alpha(\theta,\phi)$ is the angle between the direction θ, ϕ and the line joining X_1 to X_2. T_{12} denotes the distance X_2-X_1 divided by the velocity c of the incident noise.

It follows from (5) and (33) that the space correlation function of the noise is

$$\phi_n(X_1,X_2,t) = \int\int\int\int \psi_n(\theta_1,\phi_1,\theta_2,\phi_2,t+T_{12}\cos\alpha) \times \times \sin\theta_1\sin\theta_2\,d\theta_1\,d\theta_2\,d\phi_1\,d\phi_2 \quad .....(34)$$

where

$$\psi_n(\theta_1,\phi_1,\theta_2,\phi_2,t) = \lim_{T\to\infty} (1/32\pi^2 T) \int_{-T}^{T} p_n(\theta_1,\phi_1,t_1) \times \times p_n^*(\theta_2,\phi_2,t_1-t)\,dt_1 \quad .....(35)$$

The function ψ_n is the time correlation function of the component waves $p_n(\theta_1,\phi,t)$ and $p_n(\theta_2,\phi_2,t)$.

If there is no correlation between component waves incident from different directions then $\psi_n(\theta_1,\phi_1,\theta_2,\phi_2,t)$ may be written in the form

$$\psi_n(\theta_1,\phi_1,\theta_2,\phi_2,t) = \delta(\theta_1-\theta_2,\phi_1-\phi_2)\psi_n(\theta_1,\phi_1,t) \quad .....(36)$$

where $\psi_n(\theta_1,\phi_1,t)$ denotes the time autocorrelation function $\psi_n(\theta_1,\phi_1,\theta_1,\phi_1,t)$ and δ denotes the unit impulse function defined so that

$$\int\int \delta(\theta_1-\theta_2,\phi_1-\phi_2)\sin\theta_1\,d\theta_1\,d\phi_1 = 1 \quad .....(37)$$

If, furthermore, each component wave is of white noise then

$$\psi_n(\theta_1,\phi_1,t) = \psi_n(\theta_1,\phi_1)\delta(t) \quad .....(38)$$

and ψ_n is given by the expression

$$\psi_n(\theta,\phi) = \lim_{T\to\infty} (1/32\pi^2 T) \int_{-T}^{T} |p_n(\theta,\phi,t)|^2 \,\mathrm{d}t \quad \ldots\ldots(39)$$

Thus equation (34) becomes

$$\phi_n(X_1, X_2, t) = \int\int \psi_n(\theta,\phi)\delta(t+T_{12}\cos\alpha)\times \times \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi \quad \ldots\ldots(40)$$

6.1. *Noise with Cylindrical Symmetry*

In the special instance in which the noise has cylindrical symmetry about the line joining X_1 to X_2 it is convenient to choose the z axis along X_1 X_2. Then $\alpha = \theta$ and $\psi_n(\theta, \phi) = \psi_n(\theta)$ so that eqn. (40) reduces to

$$\phi_n(X_1, X_2, t) = 2\pi \int \psi_n(\theta)\delta \times \times (t+T_{12}\cos\theta)\sin\theta \,\mathrm{d}\theta \quad \ldots\ldots(41)$$

The substitution $T' = T_{12}\cos\theta$ then gives

$$\phi_n(X_1, X_2, t) = (2\pi/T_{12}) \int_{-T_{12}}^{T_{12}} \psi_n(\theta)\delta(t+T')\,\mathrm{d}T' = \begin{cases} (2\pi/T_{12})\psi_n(\theta) & \text{if } |t| < T_{12} \\ 0 & \text{if } |t| > T_{12} \end{cases} \quad \ldots\ldots(42)$$

where $\theta = \arccos(-t/T_{12})$ and is chosen within the range $0 \leq \theta \leq \pi$. Hence it follows from (6) that

$$\Phi_n(X_1, X_2, \omega) = (2\pi/T_{12}) \int_{-T_{12}}^{T_{12}} \psi_n(\theta) \times \times \exp(-\mathrm{j}\omega t)\,\mathrm{d}t \quad \ldots\ldots(43)$$

For isotropic noise that is incident in equal power from all directions in space then $\psi_n(\theta) = 1$ and eqn. (43) leads to

$$\Phi_n(X_1, X_2, \omega) = 4\pi \frac{\sin\omega T_{12}}{\omega T_{12}} \quad \ldots\ldots(44)$$

The choice of $\psi_n(\theta) = \sin^r\theta$ represents a noise field whose maximum components are from horizontal directions; the corresponding value of $\Phi_n(X_1, X_2, \omega)$ is

$$\Phi_n(X_1, X_2, \omega) = 2\pi^{3/2} \frac{\Gamma(\frac{1}{2}r+1)\mathrm{J}_{(r+1)/2}(\omega T_{12})}{(\omega T_{12}/2)^{(r+1)/2}} \quad \ldots\ldots(45)$$

6.2. *Planar Array with Planar Noise*

In the special instance in which the receiving elements and the directions of the incident noise waves are in the xy plane then

$$\psi_n(\theta,\phi) = \delta(\theta-90)\psi_n(\phi) \quad \ldots\ldots(46)$$

so that equation (40) reduces to

$$\phi_n(X_1, X_2, t) = \int_{-\pi}^{\pi} \psi_n(\phi)\delta[t+T_{12}\cos(\phi-\phi_{12})]\,\mathrm{d}\phi = \int_0^{\pi} [\psi_n(\phi_{12}+\phi)+\psi_n(\phi_{12}-\phi)] \times \times \delta(t+T_{12}\cos\phi)\,\mathrm{d}\phi \quad \ldots\ldots(47)$$

in which ϕ_{12} denotes the direction of X_2 with respect to X_1. The substitution $T' = T_{12}\cos\phi$ leads to

$$\phi_n(X_1, X_2, t) = \begin{cases} \dfrac{\psi_n(\phi_{12}+\phi)+\psi_n(\phi_{12}-\phi)}{T_{12}\sin\phi} & \text{if } |t| < T_{12} \\ 0 & \text{if } |t| > T_{12} \end{cases} \quad \ldots\ldots(48)$$

where ϕ denotes $\arccos(-t/T_{12})$ in the range $0, \pi$. It follows from (48) and (6) that

$$\Phi_n(X_1, X_2, \omega) = \int_{-T_{12}}^{T_{12}} \frac{\psi_n(\phi_{12}+\phi)+\psi_n(\phi_{12}-\phi)}{T_{12}\sin\phi} \times \times \exp(-\mathrm{j}\omega t)\,\mathrm{d}t = \int_0^{\pi} [\psi_n(\phi_{12}+\phi)+\psi_n(\phi_{12}-\phi)] \times \times \exp(\mathrm{j}\omega T_{12}\cos\phi)\,\mathrm{d}\phi \quad \ldots\ldots(49)$$

For isotropic noise $\psi_n(\phi) = 1$ and eqn. (49) leads to

$$\Phi_n(X_1, X_2, \omega) = 2\pi J_0(\omega T_{12}) \quad \ldots\ldots(50)$$

7. Form of the Signal

Suppose the signal incident upon any unit area consists of a sum of plane waves $p_s(\theta, \phi, t)$. At the time t let the signal received at X_1 be

$$p_s(X_1, t) = (1/4\pi) \int\int p_s(\theta,\phi,t)\sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi \quad \ldots\ldots(51)$$

Its spectrum is given by (16) as

$$P(X_1, \omega) = \int p_s(X_1, t)\exp(-\mathrm{j}\omega t)\,\mathrm{d}t$$

Let

$$P(\theta,\phi,\omega) = \int p_s(\theta,\phi,t)\exp(-\mathrm{j}\omega t)\,\mathrm{d}t \quad \ldots\ldots(52)$$

be the component of the spectrum received at X_1 from θ, ϕ so that

$$P(X_1, \omega) = (1/4\pi) \int\int P(\theta,\phi,\omega)\sin\theta \,\mathrm{d}\phi \,\mathrm{d}\theta$$

Then the signal received at X_2 is

$$p_s(X_2, t) = (1/4\pi) \int\int p_s(\theta,\phi,t+T_{12}\cos\alpha)\times \times \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi \quad \ldots\ldots(53)$$

and its spectrum is

$$P(X_2, \omega) = (1/4\pi) \int\int P(\theta,\phi,\omega)\exp(\mathrm{j}\omega T_{12}\cos\alpha)\times \times \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi \quad \ldots\ldots(54)$$

If the signal is from a single direction θ_c, ϕ_c then

$$P(X_2, \omega) = (1/4\pi)P(\theta_c, \phi_c, \omega)\exp(\mathrm{j}\omega T_{12}\cos\alpha)$$

For general forms of the signal, equation (54) may be rewritten by noting that

$$T_{12}\cos\alpha = T_{12}\sin\theta_{12}\cos(\phi_{12}-\phi)\sin\theta + T_{12}\cos\theta_{12}\cos\theta \quad \ldots\ldots(55)$$

and hence

$$P(X_2, \omega) = (1/4\pi) \int P(\theta,\phi,\omega)\times \times \exp(\mathrm{j}\omega T_{12}\cos\theta_{12}\cos\theta)\sin\theta \,\mathrm{d}\theta \times \times \int \exp[\mathrm{j}\omega T_{12}\sin\theta_{12}\sin\theta \times \times \cos(\phi_{12}-\phi)]\,\mathrm{d}\phi \quad \ldots\ldots(56)$$

8. Optimum Line Array when $\psi_n = \sin\theta$

Consider an array of N receivers spaced at equal intervals along the z axis as shown in Fig. 4. Suppose the noise has cylindrical symmetry about the line of the array and consists of uncorrelated white noise components such that $\psi_n(\theta) = \sin\theta$. The maximum contribution to the noise is then from horizontal directions, and from eqn. (45)

$$\Phi_n(X_1, X_2, \omega) = 2\pi^2 \frac{J_1(\omega T_{12})}{\omega T_{12}} \quad \ldots\ldots(57)$$

If the signal has symmetry about the array axis and is confined to a small angle $\Delta\theta$ as indicated in Fig. 4, then from eqn. (56)

$$P(X_2, \omega) = \tfrac{1}{2}\Delta\theta P(\omega)(\sin\theta)\exp(j\omega T_{12}\cos\theta) \quad \ldots\ldots(58)$$

After substitution for Φ_n and $P(X_2, \omega)$ in eqn. (22) the integration may be replaced by a sum over array elements. It proves convenient to denote $\omega X/c$ by Z and to define $A(Z_1, \omega)$ and $B(Z_1, \omega)$ by

$$A(Z_1, \omega) + jB(Z_1, \omega) = \frac{4\pi^2 \lambda_t R^*(X_1, \omega)\,dX}{(\Delta\theta)P(\omega)(\sin\theta)V_s^*(t)\exp(j\omega t)} \quad \ldots\ldots(59)$$

Then eqn. (22) becomes

$$\sum_{Z_1} \frac{J_1(Z_1 - Z)}{Z_1 - Z} A(Z_1, \omega) = \cos(Z\cos\theta) \quad \ldots\ldots(60a)$$

$$\sum_{Z_1} \frac{J_1(Z_1 - Z)}{Z_1 - Z} B(Z_1, \omega) = \sin(Z\cos\theta) \quad \ldots\ldots(60b)$$

Solution of the $2N$ simultaneous equations represented by (60a) and (60b) determines the parameters $A(Z_1, \omega)$ and $B(Z_1, \omega)$ of the processing networks.

The maximum value of $V_s(t)^2/P_n^2$ may be determined from eqn. (25) which may be expressed in the form

$$\frac{V_s(t)^2}{P_n^2} = \frac{\lambda_t}{2\pi V_s^*(t)} \int\int P^*(X_2, \omega) R^*(X_2, \omega) \times \exp(-j\omega t)\,d\omega\,dX_2$$

$$= \frac{(\Delta\theta)^2 \sin^2\theta}{16\pi^3} \sum_{Z_1} |P(\omega)|^2 [A(Z_1, \omega) + jB(Z_1, \omega)] \times \exp(-jZ_1\cos\theta)\,d\omega \quad \ldots\ldots(61)$$

When the response functions are optimized the individual frequency components of the signal contribute in a linear manner to $V_s(t)^2/P_n^2$ and the contribution per unit bandwidth at the frequency ω is

$$\frac{V_s(t)^2}{P_n^2} = K(\theta) \sum_{Z_1} [A(Z_1, \omega) + jB(Z_1, \omega)] \times \exp(-jZ_1\cos\theta) \quad \ldots\ldots(62)$$

where

$$K(\theta) = (\Delta\theta)^2 (\sin^2\theta) |P(\omega)|^2 / 16\pi^3 \quad \ldots\ldots(63)$$

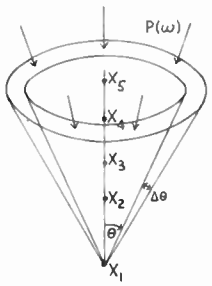


Fig. 4. Line array with $\psi_n(\theta) = \sin\theta$.

The dependence of $|V_s(t)^2/P_n^2|/K(\theta)$ upon the angle of incidence θ of the signal and the hydrophone spacing h is shown in Tables 1 and 2 for arrays consisting of five and nine receivers.

Comparison of Tables 1 and 2 indicates the advantage to be gained by the addition of four more elements

Table 1

Values of $|V_s(t)^2/P_n^2|/K$ for a five-element line array of Fig. 4 when $\psi_n(\theta) = \sin\theta$.

$90 - \theta°$	$h = \lambda/4$	$\lambda/2$	λ	2λ
0	4·75	8·22	11·81	10·58
10	5·79	8·28	10·24	9·77
20	6·85	8·76	9·42	9·81
30	6·30	9·43	9·22	9·70
40	7·19	10·69	9·42	9·85
50	13·13	14·83	9·83	9·70
60	11·29	15·52	10·64	9·88
70	33·42	16·79	11·50	10·27
80	40·37	17·13	11·78	10·56
90	42·75	17·15	11·81	10·58

Table 2

Values of $|V_s(t)^2/P_n^2|/K$ for a nine-element line array when $\psi_n(\theta) = \sin\theta$.

$90 - \theta°$	$h = \lambda/4$	$\lambda/2$	λ
0	9·55	14·61	22·47
10	9·83	15·14	17·56
20	10·51	15·42	16·83
30	11·60	16·73	16·48
40	14·45	18·96	16·76
50	16·54	22·35	17·62
60	167	30·08	19·01
70	69	36·70	20·90
80	150	38·85	22·36
90	193	39·01	22·47

to a five-hydrophone array. Consider for example an array consisting of five elements spaced at a distance of $\lambda/2$ apart. The addition of four hydrophones to form an array of spacing $\lambda/4$ leads to greatly increased signal/noise ratio for a signal incident from the direction $\theta = 0$ but produces only a slight increase in signal/noise ratio for a signal incident from the direction $\theta = 90°$. If the four additional hydrophones are placed so as to increase the array length while preserving an element spacing of $\lambda/2$ then the signal/noise ratio is approximately doubled for all directions of the incident signal.

It appears that, for the type of noise assumed in the computations leading to the results listed in Tables 1 and 2, a hydrophone spacing of $\lambda/2$ leads in general to high values of signal/noise ratio. However for an incident signal for which $\theta = 0$ it is desirable to have a smaller spacing of elements, while for a signal incident from the direction $\theta = 90°$ it is more advantageous to space the hydrophones at a distance λ apart.

9. Optimum Line Array with Plane Isotropic Noise

Now consider an array of N receivers spaced equidistantly along the y-axis and suppose that the noise is isotropic and confined to the xy plane. Then according to eqn. (50)

$$\Phi_n(X_1, X_2, \omega) = 2\pi J_0(\omega T_{12})$$

If the signal is in the xy plane and confined to a small angle $\Delta\phi$ then according to eqn. (54)

$$P(X_2, \omega) = (1/4\pi)\Delta\phi P(\omega) \exp(j\omega T_{12} \sin\phi) \quad \text{.....(64)}$$

The expressions for Φ_n and $P(X_2, \omega)$ may be substituted in to equation(22) with the integration replaced by a sum over array elements. Let $\omega X/c$ be denoted by Y and let $A(Y_1, \omega)$ and $B(Y_1, \omega)$ be defined by

$$A(Y_1,\omega)+jB(Y_1,\omega) = \frac{8\pi^2\lambda_t R^*(X_1,\omega)\,dX_1}{\Delta\phi P(\omega)V_s^*(t)\exp(j\omega t)} \quad \text{.....(65)}$$

Then eqn. (22) takes the form

$$\sum_{Y_1} J_0(Y_1 - Y)A(Y_1) = \cos(Y \sin\phi) \quad \text{.....(66a)}$$

$$\sum_{Y_1} J_0(Y_1 - Y)B(Y_1) = \sin(Y \sin\phi) \quad \text{.....(66b)}$$

The equation analogous to equation (62) is then

$$\frac{V_s(t)^2}{P_n^2} = K\sum_{Y_1}[A(Y_1,\omega)+jB(Y_1,\omega)]\times \times\exp(-jY_1 \sin\phi) \quad \text{.....(67)}$$

where

$$K = (\Delta\phi)^2|P(\omega)|^2/64\pi^2 \quad \text{.....(68)}$$

The dependence of $|V_s(t)^2/P_n^2|/K$ upon angle of the incident signal is shown in Tables 3 and 4 for arrays consisting of five and nine elements.

Table 3

Values of $|V_s(t)^2/P_n^2|/K$ for a five-element line array on the y-axis for plane isotropic noise in the xy plane.

$\phi°$	$h = \lambda/4$	$\lambda/2$	λ	2λ
0	5·41	7·46	3·00	3·39
10	4·93	7·42	5·37	5·86
20	4·52	7·01	6·41	5·73
30	5·00	6·46	6·47	3·39
40	5·04	5·98	6·49	5·62
50	4·00	4·60	5·87	6·11
60	3·38	3·39	4·81	5·57
70	4·57	2·75	3·51	4·68
80	6·74	2·57	3·03	3·50
90	7·81	2·56	3·00	3·39

Table 4

Values of $|V_s(t)^2/P_n^2|/K$ for a nine-element line array for plane isotropic noise.

$\phi°$	$h = \lambda/8$	$\lambda/4$	$\lambda/2$	λ
0	5·50	8·33	13·76	4·23
10	5·43	10·73	13·57	9·84
20	4·60	8·14	12·97	11·62
30	5·03	10·12	11·91	12·01
40	5·23	8·22	10·48	11·68
50	4·15	8·28	8·87	10·74
60			6·64	9·15
70	4·59	5·58	4·34	6·63
80	6·88	9·62	3·56	4·43
90	8·00	15·42	3·50	4·23

Comparison of Tables 3 and 4 suggests that a hydrophone spacing of $\lambda/2$ leads to high signal/noise ratios except for signals incident at small angles ϕ. For values of ϕ close to 90 deg it appears advantageous to choose an element spacing of $\lambda/4$.

The value of $A^2(Y, \omega) + B^2(Y, \omega)$ represents the power amplification of the processing network connected to the receiver situated at Y. The value of arctan (B/A) is the phase delay required from this network. For the line arrays of Tables 1–4 the elements are spaced equidistantly and the optimum processing networks produce the same phase delay between successive receivers.

The optimum form of the power amplification of the receivers varies considerably for different element spacings and directions of the incident signal. The power amplification is not, in general, a linear function of distance from the central element, nor does it taper to zero toward the ends of the line array. To

Table 5

Relative power amplification of each receiver in an optimum nine element line array for plane isotropic noise. Element spacing = $\lambda/2$.

y	$\phi = 0$	30°	60°	90°
2λ	1·75	1·50	1·00	·241
$1{\cdot}5\lambda$	2·80	1·69	·598	·154
λ	2·26	2·03	·476	·127
$0{\cdot}5\lambda$	2·64	1·88	·427	·117
0	2·30	1·71	·413	·114
$-0{\cdot}5\lambda$	2·64	1·88	·427	·117
$-\lambda$	2·26	2·03	·476	·127
$-1{\cdot}5\lambda$	2·80	1·69	·598	·154
-2λ	1·75	1·50	1·00	·241
Phase Delay	0°	90°	150°	180°
$\lvert V_s(t)^2/P_n^2\rvert/K$	13·76	11·91	6·64	3·50

illustrate these remarks the optimum power amplifications for a nine element line array are listed in Table 5 for an array spacing of $\lambda/2$ and for signals incident at $\phi = 0$, 30, 60, and 90 deg. The listed phase delay refers to the phase difference between adjacent receivers.

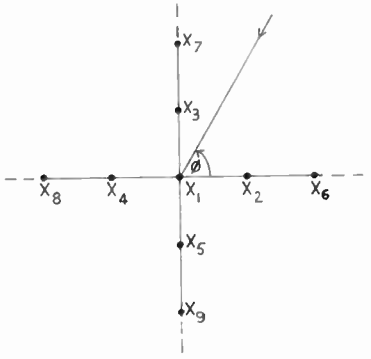


Fig. 5. Crossed array treated in Section 10.

10. Optimum Crossed Array

Suppose the receivers are equidistantly spaced along the arms of a cross as in Fig. 5. Suppose the noise to be isotropic in the plane of the receiver and suppose the signal to be from a single direction ϕ.

Each receiver is at a position X of coordinates $(x_n c/\omega, y_n c/\omega)$ where only one of x_n, y_n is different from zero. It follows from eqns. (50) and (54) that

$$\Phi_{nm} = 2\pi J_0[(x_n - x_m)^2 + (y_n - y_m)^2]^{\frac{1}{2}} \quad \text{.....(69)}$$

and

$$P(X,\omega) = (1/4\pi)(\Delta\phi)P(\omega) \times \times \exp(\mathrm{j}x\cos\phi + \mathrm{j}y\sin\phi) \quad \text{.....(70)}$$

Then eqn. (22) is equivalent to

$$\sum_m J_0[(x_n - x_m)^2 + (y_n - y_m)^2]^{\frac{1}{2}} A(x_m, y_m) = \cos(x_n\cos\phi + y_n\sin\phi) \quad \text{.....(71a)}$$

$$\sum_m J_0[(x_n - x_m)^2 + (y_n - y_m)^2]^{\frac{1}{2}} B(x_m, y_m) = \sin(x_n\cos\phi + y_n\sin\phi) \quad \text{.....(71b)}$$

where

$$A + \mathrm{j}B = \frac{8\pi^2\lambda_t R^*(X,\omega)\,\mathrm{d}X}{(\Delta\phi)P(\omega)V_s^*(t)\exp(\mathrm{j}\omega t)} \quad \text{.....(72)}$$

The equation corresponding to (62) is

$$\frac{V_s(t)^2}{P_n^2} = K\sum_m [A(x_m,\omega) + \mathrm{j}B(x_m,\omega)] \times \times \exp(-\mathrm{j}x_m\cos\phi - \mathrm{j}y_m\sin\phi) \quad \text{.....(73)}$$

where K is as in eqn. (68).

For arrays that have a total of five and nine elements the dependence of $\lvert V_s(t)^2/P^2\rvert K$ upon angle of incidence is shown in Tables 6 and 7.

Comparison of Tables 3 and 6 and also of Tables 4 and 7 leads to the conclusion that, for many of the listed hydrophone spacings and angles of incidence, a cross array does not offer any advantage over the line array from the point of view of increasing the

Table 6

Values of $\lvert V_s(t)^2/P_n^2\rvert/K$ for a five-element cross array of Fig. 5 with plane isotropic noise in plane of array.

ϕ°	$h = \lambda/4$	$\lambda/2$	λ	2λ
0	6·42	5·14	3·89	4·15
10	5·51	4·69	4·34	4·57
20	4·15	6·14	4·69	4·99
30	4·20	4·94	5·20	4·60
40	5·15	6·03	6·31	6·25
45	5·36	6·23	6·54	6·76

Table 7

Values of $\lvert V_s(t)^2/P_n^2\rvert/K$ for a nine-element cross array of Fig. 5 with plane isotropic noise.

ϕ°	$h = \lambda/8$	$\lambda/4$	$\lambda/2$	λ
0	8·71	11·58	6·56	4·74
10	7·81	8·60	7·61	8·11
20	6·09	7·31	7·97	8·95
30	6·03	9·65	8·01	9·21
40	6·58	7·16	8·16	10·82
45	6·75	6·36	8·13	11·84

Table 8

Values of $|V_s(t)^2/P_n^2|/K$ for a five-element L-array for plane isotropic noise.

$\phi°$	$h = \lambda/4$	$\lambda/2$	λ
0	4·98	4·79	38·63
10	1·60	3·54	18·58
20	3·82	·20	20·94
30	13·25	·10	3·83
40	23·10	·74	15·94
45	31·47	2·37	30·91

signal/noise ratio. However, in view of the symmetry about $\phi = 45$ deg, the signal/noise ratio for the cross array tends to be less dependent upon the direction of the incident signal. This is particularly true when the hydrophone spacing is $\lambda/2$.

11. Multiplicative Optimum Array

Consider a multiplicative array in which the receivers are arranged along an L-shaped array consisting of two of the four arms of the cross shown in Fig. 5. Let the receivers X_1, X_2, X_6, . . . correspond to the dX of Section 5, and let the receivers X_1, X_3, X_7, . . . correspond to the dX'. The central element is thus common to both portions of the multiplicative system.

In the notation of Section 9, eqn. (31a) takes the form

$$\sum_m J_0(x_n^2+y_m^2)^{\frac{1}{2}} A_m = \cos(x_n \cos\phi) \quad \text{.....(74a)}$$

$$\sum_m J_0(x_n^2+y_m^2)^{\frac{1}{2}} B_m = \sin(x_n \cos\phi) \quad \text{.....(74b)}$$

where the x_n correspond to the elements 1, 2, 6, . . . and the y_m correspond to the elements 1, 3, 7, ... Equation (31b) assumes the form

$$\frac{V_s(t)^2}{P_n^2} = K \sum_m (A_m + jB_m) \exp(-j y_m \sin\phi) \quad \text{.....(75)}$$

For multiplicative arrays of five elements the dependence of $|V_s(t)^2/P_n^2|/K$ is shown in Table 8.

Comparison of Table 8 with Tables 6 and 7 indicates that the optimum signal/noise ratio for the multiplicative array is much more dependent upon the direction of the signal and upon the spacing of the receivers. By suitable choice of receiver spacing the multiplicative array may be designed to produce very large signal/noise ratios.

12. Concluding Remarks

The theory outlined in Sections 2–5 has been applied in Sections 8–11 to determine optimum line, cross, and multiplicative arrays. The signal and noise are as described in Sections 6 and 7.

It should be emphasized that throughout the present paper the criterion for optimum is maximization of signal/noise ratio. The ratios listed in the Tables do not describe directivity patterns in the usual sense since the required response $R(X, \omega)$ of each processing network is a function of the signal direction. Appropriate variation of the $R(X, \omega)$ steers the array to remain optimum as the signal direction varies.

The optimum signal/noise ratios are dependent upon the cross-correlation function of the noise as well as upon the signal. An incorrect assumption concerning the form of the noise will, of course, result in a false estimate of the optimum signal/noise ratio. It is believed that some of the high signal/noise ratios predicted in the present paper are very sensitive to small changes in the assumed form of the noise. When using an analysis of the type described it is therefore advisable either to use a measured value of the cross-correlation function of the noise or else to repeat the calculations for a number of possible forms of the function.

In spite of the qualifications outlined above, it is believed that the theoretical analysis is useful in providing an optimum system with which to compare existing ones.

13. References

1. W. W. Hansen and J. R. Woodward, "A new principle in directional array design", *Proc. Inst. Radio Engrs*, **26**, p. 333, 1938.
2. S. A. Schelkunoff, "A mathematical theory of linear arrays", *Bell Syst. Tech. J.*, **22**, p. 80, 1943.
3. C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level", *Proc. I.R.E.*, **34**, p. 335, 1946.
4. H. J. Riblet and C. L. Dolph, Discussion on "A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level", *Proc. I.R.E.*, **35**, p. 489, 1947.
5. G. Sinclair and F. V. Cairns, "Optimum patterns for arrays of non-isotropic sources", *Trans. I.R.E. on Antennas and Propagation*, AP-1, p. 50, 1952.
6. R. H. Duhamel, "Optimum patterns for endfire arrays", *Proc. I.R.E.*, **41**, p. 652, 1953.
7. D. Barbiere, "A method for calculating the current distribution of Tschebyscheff arrays", *Proc. I.R.E.*, **40**, p. 78 and p. 991, 1952.
8. R. J. Stegen, "Excitation coefficients and beamwidths of Tschebyscheff Arrays", *Proc. I.R.E.*, **41**, p. 1671, 1953.
9. D. R. Rhodes, "The optimum linear array for a single main beam", *Proc. I.R.E.*, **41**, p. 793, 1953.
10. R. L. Pritchard, "Optimum directivity patterns for linear point arrays", *J. Acoust. Soc. Amer.*, **25**, p. 879, 1953.
11. R. L. Pritchard, "Approximate calculation of the directivity factor of linear point arrays", *J. Acoust. Soc. Amer.*, **25**, p. 1010, 1953.
12. R. L. Pritchard, "Maximum directivity index of a linear point array", *J. Acoust. Soc. Amer.*, **26**, p. 1034, 1954.

13. D. G. Tucker, "Some aspects of the design of strip arrays", *Acustica*, **6**, p. 403, 1956.
14. A. Berman and C. S. Clay, "Theory of time-averaged-product arrays", *J. Acoust. Soc. Amer.*, **29**, p. 805, 1957.
15. D. C. Fakley, "Comparison between the performance of a time-averaged product array and the intraclass correlator", *J. Acoust. Soc. Amer.*, **31**, p. 1307, 1959.
16. V. G. Welsby and D. G. Tucker, "Multiplicative receiving arrays", *J.Brit.I.R.E.*, **19**, p. 369, 1959.
17. V. G. Welsby, "Multiplicative receiving arrays: the angular resolution of targets in sonar system with electronic scanning", *J. Brit.I.R.E.*, **22**, p. 5, 1961.
18. D. G. Tucker, "Multiplicative arrays in radio astronomy and sonar", *The Radio and Electronic Engineer*, **25**, p. 113, 1963.
19. V. G. Welsby, "The signal/noise gain of ideal receiving arrays", *Proc. Instn Elect. Engrs*, **109**, Part C, p. 108, 1962.
20. D. G. Tucker, "Signal/noise performance of superdirective arrays", *Acustica*, **8**, p. 112, 1958.
21. D. G. Tucker, "The signal/noise performance of electro-acoustic strip arrays", *Acustica*, **8**, p. 53, 1958.
22. H. S. Heaps, "General theory for the synthesis of hydrophone arrays", *J. Acoust. Soc. Amer.*, **32**, p. 356, 1960.

14. Appendix

In view of the importance of eqn. (10) it appears desirable to include an elementary proof of its validity.

For a discrete set of receivers each occupying a volume $\mathrm{d}X$ then eqn. (9) may be written in the form

$$P_s^2 = (1/2\pi)(\mathrm{d}X)^2 \int \sum_{X_1} \sum_{X_2} R(X_1, \omega) R^*(X_2, \omega) \times \times \Phi_s(X_1, X_2 \omega) \mathrm{d}\omega$$

With arbitrarily small error this latter equation may be written in the equivalent form

$$P_s^2 = (1/2\pi)(\mathrm{d}X)^2 \, \mathrm{d}\omega \sum_n \sum_m \sum_q R_{nq} R_{mq}^* \Phi_s^{nmq}$$

with suitable definition of R_{nq} and Φ_s^{nmq}. Similarly

$$P_n^2 = (1/2\pi)(\mathrm{d}X)^2 \, \mathrm{d}\omega \sum \sum \sum R_{nq} R_{mq}^* \Phi_n^{nmq}$$

and hence

$$\frac{P_s^2}{P_n^2} = \frac{\sum \sum \sum R_{nq} R_{mq}^* \Phi_s^{nmq}}{\sum \sum \sum R_{nq} R_{mq}^* \Phi_n^{nmq}}$$

The ratio P_s^2/P_n^2 may be regarded as a function of a set of independent complex numbers R_{nq} each of which contains a real and imaginary component. Instead of regarding P_s^2/P_n^2 as a function of the real and imaginary parts of each R_{nq} it is permissible to regard P_s^2/P_n^2 as a function of independent quantities R_{nq} and R_{nq}^*. Then P_s^2/P_n^2 is stationary when the R_{nq} are chosen so that for each n, q

$$(\partial/\partial R_{nq})(P_s^2/P_n^2) = 0$$

hence

$$P_n^2 \sum_m R_{mq}^* \Phi_s^{nmq} - P_s^2 \sum_m R_{mq}^* \Phi_n^{nmq} = 0$$

and thus for each n, q

$$\sum_m (\Phi_s^{nmq} - \lambda_m \Phi_n^{nmq}) R_{mq}^* = 0 \qquad \ldots\ldots(76)$$

where λ_m denotes the stationary value of P_s^2/P_n^2. Equation (76) is equivalent to the following equation valid for X_1 and ω

$$\int [\Phi_s(X_1, X_2, \omega) - \lambda_m \Phi_n(X_1, X_2, \omega)] \times \times R^*(X_2, \omega) \, \mathrm{d}X_2 = 0 \qquad \ldots\ldots(77)$$

and this is eqn. (10).

Equation (77) is valid whenever P_s^2/P_n^2 is stationary with respect to $R(X_1, \omega)$ the stationary values being those values of λ_m for which eqn. (77) admits a solution for $R^*(X_2, \omega)$. The maximum value of P_s^2/P_n^2 is included among the stationary values and is clearly the maximum value of λ_m for which eqn. (77) has a solution.

Manuscript received by the Institution on 1st April 1964. (Paper No. 960.)

DISCUSSION

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

Mr. L. C. Walters: The author has indicated that performance of 'optimized' arrays is extremely sensitive to small variations in parameter values. This is not surprising since it would appear that what Professor Heaps is doing is to produce synthetically a supergain array. Such arrays are well-known to be extraordinarily critical when large supergain factors are desired and the criticality may make them of little practical value for most applications.

Professor Heaps (*in reply*)**:** Extreme sensitivity results when there is considerable correlation between the noise at different receivers. For example it is apparent from eqn. (13) that large signal/noise ratios result when n is approximately $\pm$ 1. The form of noise described by eqns. (50) or (57) leads to significant correlation for certain spacings of receivers. It is believed that other forms of noise field might be more realistic.

U.D.C. 621.396.677

The Possibility of an Interaction Anomaly in Acoustic Receiving Arrays and Radio Arrays

By

J. S. M. RUSBY, Ph.D.†

Summary: The conditions whereby interaction anomalies which occur in transmitting arrays could arise in acoustic or radio receiving arrays are established. It is concluded that the effect will only occur in special cases: for close packed resonant acoustic arrays and for radio arrays with thick dipoles arrayed asymmetrically.

1. Introduction

In November 1959 it was found that if acoustically small low-loss projectors were worked in close-packed arrays, large variations in both the phase and amplitude of the radiating face velocities of neighbouring elements occurred near resonance.[1] This anomalous behaviour is due to the high mutual radiation impedances involved which at resonance control the steady-state radiating face velocities of the elements since their mechanical impedances are then very small. In projector arrays, often driven at high powers, the anomaly could lead to projector or generator breakdown due to overstressing of the mechanical or electrical circuits. It is the object of this paper to continue the investigation in order to examine whether a similar anomaly could exist in acoustic receiving arrays or in the allied field of radio arrays.

2. Acoustic Receiving Arrays

Because of the reciprocal relationship existing between transmitting and receiving arrays it is likely that the anomaly found in projector arrays has a counterpart in receiver arrays. This phenomenon could only occur if certain requirements were met, identical to those in a projector array, namely that

(*a*) the diaphragms of the elements are acoustically small, say less than $\lambda/3$,

(*b*) the inter-element spacing was less than $\lambda/2$,

(*c*) low-loss transducer elements were used, working near resonance,

(*d*) the array of configuration was such that there was no unique value of transducer radiation impedance.

If anomalous behaviour does occur in a receiver array then it will also be dependent on the method of electrical connection between elements and on the output load of the system. For piezoelectric elements the worst conditions would occur with parallel connections and an output load which was small compared with the modulus of the equivalent circuit impedance of the array.

† National Institute of Oceanography, Wormley, Godalming, Surrey.

The simplest method of discovering if such an anomaly exists is to take a reversible array which exhibits the phenomenon on transmission and to use it as a receiving array. This has been done theoretically using the 5-element cruciform array described in reference 2 and evaluating its performance as a receiver. The elements are inertia-driven ADP piezoelectric transducers with rigid glass diaphragms $\lambda/6$ square at resonance (18 kc/s). Their configuration is shown in Fig. 1. In the analysis given in the Appendix, the mechanical equivalent circuits, which are lossless, are connected in parallel working into a constant electrical output load.

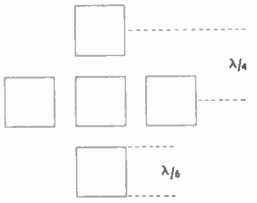


Fig. 1. Five-element cruciform array.

The results for an identical free field pressure p_{ff} on each diaphragm (i.e. normal incidence) shows that the output voltage E generated across the load R_L contains anomalous phase information concerning the received signal. The sudden alteration in phase produced at 17·7 kc/s is clearly seen in Fig. 2. In Fig. 3 the response of the receiving array, 10 log $(|E|/p_{ff})^2$ has been plotted as a function of output load. Here again the anomaly is apparent, increasing in severity as the load resistance decreases. The three curves represent cases where the output load is high compared with the resistive component of the total

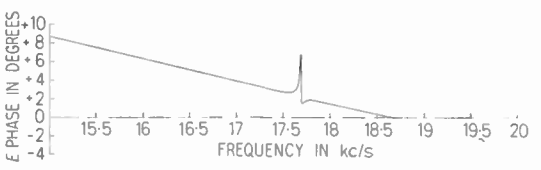


Fig. 2. Phase of output voltage with reference to free field pressure.

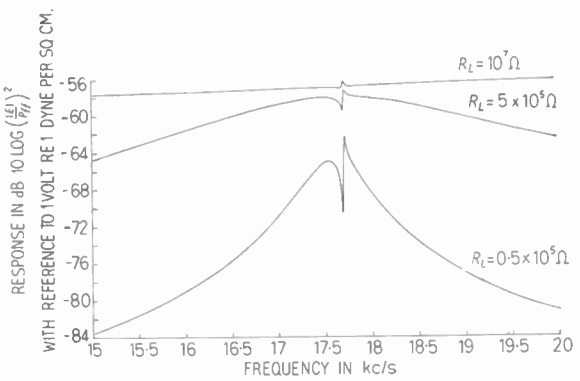


Fig. 3. Response of the receiving array as a function of output load.

impedance Z_0 of the array ($R_L = 10^7\,\Omega$), comparable with it ($R_L = 5 \times 10^5\,\Omega$) and low compared with it ($R_L = 5 \times 10^4\,\Omega$). Computations were also carried out on the ratio between the signal voltage and the thermal noise voltage generated across the load. In Fig. 4 a plot of this computed signal/noise ratio, 10 log $(|E|/E_n)^2$, is given with reference to a free field pressure of 1 dyne/cm^2 and a load of $10^7\,\Omega$. A 4-dB drop has occurred at 17·7 kc/s due to the anomaly. The thermal noise voltage generated across the load, R_L, is given by $E_n = \sqrt{(4KTR)}$ for a bandwidth of 1 c/s, where K is the Boltzmann constant, T is the absolute temperature and R is the real component of impedance across which the thermal agitation is developed, i.e. Re$[R_L Z_0/(R_L + Z_0)]$. Further calculations also showed that the response of the array to far-field signals off the axis (i.e. array gain) was affected by the interaction.

Summarizing, we can say that the analysis demonstrated that under certain conditions it is possible for a receiving array to exhibit anomalous behaviour as a result of inter-element coupling. However, it should be remembered that few practical receiver systems have a specification which will encourage the generation of the anomaly. This is because they are usually broad-band devices where power efficiency is unimportant so they operate below rather than at resonance. There is also no necessity for the elements to be close-packed

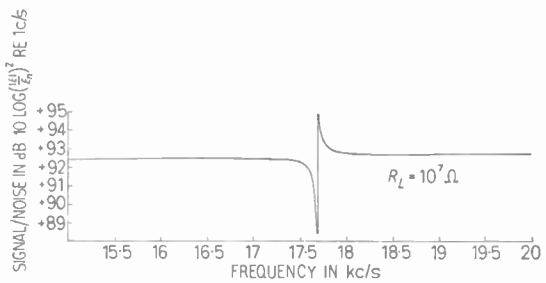


Fig. 4. Signal/noise ratio with reference to a free-field pressure of 1 dyne/cm^2 and a load of $10^7\,\Omega$.

in the array. If, however, a close-packed resonant array should be used, perhaps in the case of a reversible system acting as both transmitter and receiver, then the most important effect of the anomaly will be to produce large phase changes near resonance resulting in reduced gain and also in bearing errors in arrays scanned by phase-shifting networks. This last result arises because the existing phase errors will be increased by the insertion of reactive circuits.

3. Radio Arrays

Because of the analogy between acoustic and radio arrays it is important to discover whether the anomaly has produced comparable effects in radio. For the anomaly to be present certain requirements would have to be met:

(*a*) resonant antenna elements would have to be used;

(*b*) the mutual-radiation impedance between elements would have to be comparable with the self-radiation impedance; and

(*c*) feeder lines connecting the main transmission line to individual elements would have to be of low loss.

Of these requirements (*a*) and (*c*) are frequently met but not (*b*). In curtain arrays the active dipoles are usually placed about $\lambda/2$ apart so that the mutual radiation impedance in h.f. and v.h.f. arrays is kept small.[3] For example, in a parallel array of four thin half-wavelength active dipoles spaced $\lambda/2$ apart and fed by identical currents, the radiation resistance of the centre two elements is reduced from 74 to 52 Ω and the outside elements from 74 to 64 Ω as a result of mutual impedance.[4] This is a comparatively minor effect which is unlikely to generate an erratic distribution of currents near resonance.

The effectiveness of 'parasitic' elements gives ample evidence of the existence of mutual impedance effects for smaller spacings, in the range $\lambda/10$ to $\lambda/4$, but such spacings are seldom used between active elements at resonance.

However, there is one particular case when the mutual-radiation impedance of active elements can be large by comparison with the self-radiation impedance; in curtain arrays of thick dipoles whose length to breadth ratio is 10 : 1 or less. Here the increase in thickness reduces the self- and increases the mutual-radiation resistance very severely. These are practical arrays for broad-band operation because the increase in dipole diameter produces an input impedance with only a small reactance change as a function of frequency. The writer has been told of recent cases where the anomaly has arisen because of these large dipole widths. In these cases the arrays were initially asymmetric in configuration and were cured by adopting a

symmetric design showing that interaction was the cause of the trouble. As they were receiving arrays the anomalous behaviour was apparent from bearing discrepancies produced by large phase changes in the received signals from the elements.

The writer has not been able to find any theoretical work on anomalous interaction in radio arrays although it is clear from the literature that the inherent difficulties of working arrays with high inter-element coupling are understood. This is especially emphasized in a paper by Kurtz and Elliott[5] showing how the interaction between antennae can lead to bearing errors in arrays scanned by phase-shifting networks.

4. Conclusions

The analysis has shown that under certain conditions it is possible for an acoustic receiving array to exhibit anomalous behaviour as a result of inter-element coupling. However, in practice few receiver systems are likely to suffer from these effects. This is because they are usually broad-band devices where power efficiency is unimportant so they operate below rather than at resonance. There is also no necessity for the elements to be close-packed in the array. If, however, a close-packed resonant array should be used, perhaps in the case of a reversible system acting as both transmitter and receiver, then the most important effect of the anomaly will be to produce phase changes near resonance resulting in bearing errors and system loss.

It is apparent that the manifestation of anomalous interaction in radio arrays is also likely to be rare since the requirements for generating it are infrequently met. Practical arrays usually contain their active elements spaced at least $\lambda/2$ apart where the mutual-radiation impedance is small. When thick dipoles have been used for broad-band operation the anomaly has arisen and was cured by re-positioning the elements symmetrically.

5. References

1. J. S. M. Rusby, "Measurements of the total acoustic radiation impedance of rigid pistons in an array", *Nature*, **186**, pp. 144–5, 9th April 1960.
2. J. S. M. Rusby, "Investigation of a mutual impedance anomaly between sound projectors mounted in an array", *Acustica*, **14**, pp. 127–37, March 1964.
3. E. V. D. Glazier and H. R. L. Lamont, "The Services Textbook of Radio, Vol. 5: Transmission and Propagation", p. 306 (H.M.S.O., London, 1958).
4. E. B. Moullin, "Radio Aerials", p. 104 (Oxford University Press, 1950).
5. L. A. Kurtz and R. S. Elliott, "Systematic errors caused by the scanning of antenna arrays: phase shifters in the branch lines", *I.R.E. Trans. on Antennas and Propagation*, **AP-4**, p. 619, 1956.

6. Appendix

The performance of the 5-projector cruciform array described in reference 2 was evaluated as a receiver using the equivalent circuit for each element shown in Fig. 5.

If each transducer element in the cruciform array is treated as a linear passive four-terminal network we can write the following four equations,

for the central projector

$$\left.\begin{aligned} E &= \alpha f_1 + \beta v_1 \\ i_1 &= \gamma f_1 + \delta v_1 \end{aligned}\right\} \quad \ldots\ldots(1)$$

and for an outside projector

$$\left.\begin{aligned} E &= \alpha f_2 + \beta v_2 \\ i_2 &= \gamma f_2 + \delta v_2 \end{aligned}\right\} \quad \ldots\ldots(2)$$

where, in the parallel connection shown in Fig. 5, E is the common output voltage. In these equations

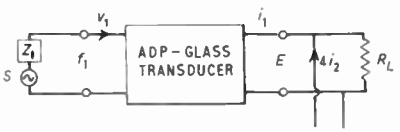


Fig. 5. Equivalent circuit for array.

α, β, γ, δ are transduction coefficients describing the mechanical circuit of the ADP-glass transducers, f and v are the force and velocity at the face of a transducer, and i is the output current. Now from Fig. 5, if Z_1 and Z_2 are the total radiation impedances of the central and an outside transducer,

$$f_1 = s - Z_1 v_1 = s - (Z_{11} v_1 + 4 Z_{12} v_2)$$

$$f_2 = s - Z_2 v_2 = s - [Z_{12} v_1 + (Z_{11} + 2Z_{23} + Z_{24}) v_2]$$

where $s = p_{ff} A D$. Here p_{ff} is the free field pressure of a normally incident plane wave, A is the area of a transducer face, and D is the diffraction constant, i.e. the ratio of blocked diaphragm pressure to the free field pressure. Substituting into eqns. (1) we get

$$\left.\begin{aligned} E &= \alpha s + (\beta - \alpha Z_{11}) v_1 - 4\alpha Z_{12} v_2 \\ i_1 &= \gamma s + (\delta - \gamma Z_{11}) v_1 - 4\gamma Z_{12} v_2 \end{aligned}\right\} \quad \ldots\ldots(3)$$

and in eqns. (2)

$$\left.\begin{aligned} E &= \alpha s - \alpha Z_{12} v_1 + [\beta - \alpha(Z_{11} + 2Z_{23} + Z_{24})] v_2 \\ i_2 &= \gamma s - \gamma Z_{12} v_1 + [\delta - \gamma(Z_{11} + 2Z_{23} + Z_{24})] v_2 \end{aligned}\right\} \quad \ldots(4)$$

From eqns. (3) and (4)

$$\frac{v_1}{v_2} = \mu = 1 + \frac{3Z_{12} - 2Z_{23} - Z_{24}}{(\beta/\alpha) - Z_{11} + Z_{12}}$$

and

$$i = i_1 + 4i_2 = 5\gamma s + \{\mu[\delta - \gamma(Z_{11} + 4Z_{12})] + 4[\delta - \gamma(Z_{11} + Z_{12} + 2Z_{23} + Z_{24})]\} v_2 \quad \ldots\ldots(5)$$

Now from eqns. (3)

$$E = \alpha s + \{\mu(\beta - \alpha Z_{11}) - 4\alpha Z_{12}\} v_2$$

hence using eqn. (5)

$$E = s\left\{\alpha - \frac{5\gamma[\mu(\beta-\alpha Z_{11})-4\alpha Z_{12}]}{\mu[\delta-\gamma(Z_{11}+4Z_{12})]+4[\delta-\gamma(Z_{12}+Z_{11}+2Z_{23}+Z_{24})]}\right\} - \left\{\frac{[4\alpha Z_{12}-\mu(\beta-\alpha Z_{11})]}{\mu[\delta-\gamma(Z_{11}+4Z_{12})]+4[\delta-\gamma(Z_{12}+Z_{11}+2Z_{23}+Z_{24})]}\right\} i$$

In this last equation the receiving array is reduced to the simple equivalent circuit of Fig. 6:

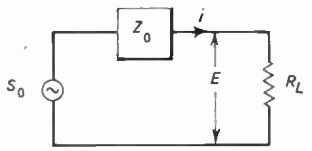


Fig. 6. Reduction of array equivalent circuit.

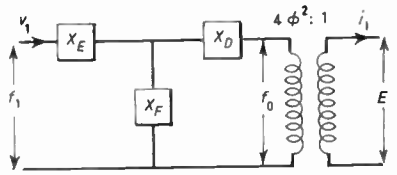


Fig. 7. Equivalent circuit used to replace the unknown four-terminal network of the ADP-glass transducer.

where the equivalent source level is

$$s_0 = s\left\{\alpha - \frac{5\gamma[\mu(\beta-\alpha Z_{11})-4\alpha Z_{12}]}{\mu[\delta-\gamma(Z_{11}+4Z_{12})]+4[\delta-\gamma(Z_{12}+Z_{11}+2Z_{23}+Z_{24})]}\right\} \quad \ldots\ldots(6)$$

and the equivalent impedance is

$$Z_0 = \frac{[4\alpha Z_{12}-\mu(\beta-\alpha Z_{11})]}{\mu[\delta-\gamma(Z_{11}+4Z_{12})]+4[\delta-\gamma(Z_{12}+Z_{11}+2Z_{23}+Z_{24})]} \quad \ldots\ldots(7)$$

Using eqns. (6) and (7) it is possible to compute the performance of the cruciform array. However, to do this the unknown transduction coefficients α, β, γ, δ must be given in terms of the known block mechanical impedances X_D, X_E, X_F of an ADP-glass transducer. In Fig. 7 the equivalent circuit employed in reference 2 has been used to replace the unknown four-terminal network, and from simple circuit analysis it can be shown that

$$\left.\begin{aligned} E &= \left(\frac{X_F+X_D}{2\phi X_F}\right) f_1 + \frac{2\phi}{X_F} \times \\ &\quad \times \left(\frac{X_F}{4\phi^2} - \frac{(X_E+X_F)(X_D+X_F)}{4\phi^2}\right) v_1 \\ i_1 &= -\frac{2\phi}{X_F} f_1 + 2\phi\left(\frac{X_E+X_F}{X_F}\right) v_1 \end{aligned}\right\} \quad \ldots\ldots(8)$$

Comparing eqns. (8) with (1) it is clear that

$$\alpha = +\frac{1}{2\phi}\left(1+\frac{X_D}{X_F}\right)$$

$$\beta = -\frac{1}{2\phi}\left(X_D+X_E+\frac{X_E X_D}{X_F}\right)$$

$$\gamma = -\frac{2\phi}{X_F}$$

$$\delta = +2\phi\left(1+\frac{X_E}{X_F}\right)$$

It is now possible to compute the performance of the cruciform array since all the quantities in eqns. (6) and (7) are known. A *Pegasus* computer was programmed to calculate the equivalent source level, s_0, and the response of the array, 10 log $(|E|/p_{ff})^2$, as a function of frequency near resonance. Computations were also made of the thermal noise voltage generated across the terminals of the load, R_L, due to thermal agitation in the electrical, mechanical and acoustic parts of the system. This voltage is given by

$$E_n = \sqrt{4KTR} \quad \text{for a bandwidth of 1 c/s}$$

where K is Boltzmann's constant and equal to $1{\cdot}37 \times 10^{-23}$ joule per degree Kelvin, T is the absolute temperature in degrees Kelvin, and R is the real component of impedance across which the thermal agitation is developed, i.e. Re$[R_L Z_0(R_L+Z_0)]$.

Manuscript first received by the Institution on 20th July 1964 and in final form on 17th November 1964. (Paper No. 961.)

U.D.C. 621.396.677

Statistical Optimization of Antenna Processing Systems

By

G. O. YOUNG, Ph.D.†

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

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

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

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

1. Introduction

Classical antenna design assumes non-random sinusoidal signals; however, in real environments, the signals of interest are always corrupted by noise and are rarely truly sinusoidal. In fact, very often the signal itself is a random variable. These conditions clearly suggest that statistical procedures must be incorporated in the design and optimization of antenna processing systems. This fact has been recognized for some time by system designers who employ statistical design procedures and optimization criteria.

The different approaches used by the system engineer and antenna engineer lead to a dichotomy which produces a 'mismatch' at the 'interface' between the antenna and the system. The antenna and the system processor are each designed and optimized independently and moreover based on completely different criteria. In order to obtain the optimum system performance it is quite apparent that the overall system must be optimized jointly and based on a uniform set of criteria and objectives. The proper approach is thus to optimize the overall system based on statistical criteria.

In order to permit the optimization of the overall system from a statistical standpoint, this paper treats the antenna as an information processing device. By representing the antenna as a spatial (as well as temporal) frequency filter, information-theoretic concepts can be applied so as to optimize the system.

† Hughes Aircraft Company, Culver City, California.

2. The Information-Theoretic Criterion

An appropriate criterion is one which minimizes the expected value of an appropriately weighted sum of the powers of the error which guarantees a good fit of the actual output to the desired output for non-Gaussian as well as Gaussian distributions.

Two specific criteria are developed. The first criterion states that the total output useful entropy and the useful entropy per output sample be jointly maximized. If in addition to providing the maximum information at the output an information retrieval requirement is imposed, the optimum system should yield the least possible error. Consequently, a second criterion is proposed which minimizes the error in an information theoretic sense. The criterion is the minimization of the total useful entropy of the system error within the filter bandwidth.

A derivation of the criteria now follows. A very useful criterion which takes into account fluctuations about the mean is minimization of the ensemble average of the square of the difference between the output and the desired output of the antenna or optical system. If the input and output distributions are both Gaussian, minimizing the mean and mean square

value of the difference between the desired output and the actual output forces the probability distribution of the output to be as similar as possible to the probability distribution of the desired signal. When the two probability distributions are as alike as possible, the actual output and the desired output are as alike statistically as is possible. Now if the signal input is not Gaussian, then the probability distribution of the input is not completely defined by the first and second moments. In general, all of the higher order moments or ensemble averages must be known to specify completely such a distribution. Thus, in this case, not only the mean square value of the difference between the output and the desired signal input should be minimized, but the mean cubed value of this quantity, the mean fourth power, etc. If all of these moments up to and including the nth, where n approaches infinity, are minimized, the probability distribution of the output is as similar as is possible to the probability distribution of the desired signal input (except in certain rare distributions not commonly encountered in practice).

A reasonable criterion for optimization is, then, that some weighted combination of the moments $\overline{(y-s)^m}$, $m = 1, 2, \ldots, \infty$, be minimized, where the desired signal output is s and the observed output is y.

The importance of a particular moment depends upon the general shape of the probability distribution defining the moment. Therefore, the appropriate weighting is usually specified by the distribution itself. Let us now search for a suitable criterion which involves the probability distribution of the random variables rather than the moments thereof. The purpose of this is to find some operation on the probability distribution having the desired moments which will minimize all of the moments with the proper weighting. Thus, the minimization of $\overline{(y-s)^m}$ may often be accomplished by simply maximizing $p(y|s)$.

Write $\overline{(y-s)^m}$ as

$$\overline{(y-s)^m} = \int_{-\infty}^{\infty}\int (y-s)^m p(y,s)\,dy\,ds$$

$$= \int_{-\infty}^{\infty} p(s)\,ds \int_{-\infty}^{\infty} (y-s)^m p(y|s)\,dy \quad \ldots\ldots(1)$$

Now it turns out that the more similar y and s become, the larger $p(y|s)$ becomes in the vicinity of $y = s$.

Indeed, if $y = s$

$$p(y|s) = \delta(y-s) \quad \ldots\ldots(2)$$

When eqn. (2) is substituted in eqn. (1), the result is

$$\overline{(y-s)^m} = 0 \quad \ldots\ldots(3)$$

If y and s are statistically independent, then

$$p(y|s) = p(y) \quad \ldots\ldots(4)$$

Under these circumstances, $\overline{(y-s)^m}$ becomes

$$\overline{(y-s)^m} = \sum_{r=0}^{m} \binom{m}{r} (-1)^{m-r}\, \overline{y^r}\, \overline{s^{m-r}} \quad \ldots\ldots(5)$$

In many practical cases the desired signal s will be Gaussian with 0 mean, even though the observed output is very often not Gaussian. Even if the distribution of s is not Gaussian, it is most likely to be symmetric about zero. Under such circumstances,

$$\overline{s^{m-r}} = 0 \qquad \text{for } m-r \text{ odd}$$

The terms in the sum in eqn. (5) are negative for $m-r$ odd. It therefore follows that if y and s are statistically independent, and s has a symmetric probability distribution about a zero mean, all of the negative terms in the sum in eqn. (5) vanish and mth moment of the difference between y and s is maximized. Thus, in general, the closer y and s are related (a more proper statistical term would be the more closely y and s are correlated) the larger $p(y|s)$ gets in the vicinity of s, and the smaller $\overline{(y-s)^m}$ becomes. The criterion of the minimization of a weighted combination of terms of the form $\overline{(y-s)^m}$ could therefore be replaced by the criterion of maximizing $p(y|s)$ at s.

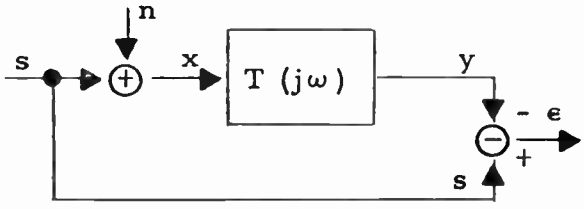


Fig. 1. General linear system.

Observe that $p(y|s)$ will in most cases be a complicated exponential function. The form of the function in general will be simplified if the logarithm to the base e of $p(y|s)$ is taken, so that instead of maximizing the exponential function, the exponent is maximized. Since maximizing the exponent maximizes the exponential function, the same result is achieved. Let us therefore require that $\ln p(y|s)$ be maximized. Now it would be convenient to normalize the expression $\ln p(y|s)$ in such a way that when y and s are independent, the measure of quality is zero, without affecting the fact that when y is equal to s, the measure is infinite (see eqn. (2)). This normalization is accomplished by dividing $p(y|s)$ by $p(y)$, so that the normalized expression to be maximized is $\ln \left[\frac{p(y|s)}{p(y)}\right]$.

$\ln \left[\frac{p(y|s)}{p(y)}\right]$ is maximum when $y = s$ and is zero when y is independent of s. This function is a random variable in the sense that both y and s are random

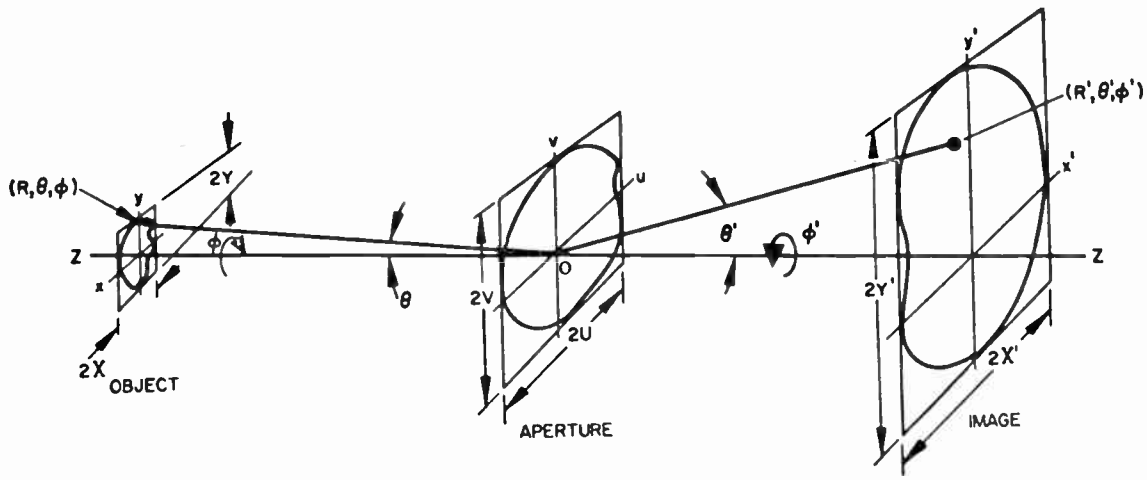


Fig. 2. Spatial geometry.

variables and that a function of a random variable is a random variable. Since a random variable is not a desirable measure of quality, an appropriate measure is found by taking the expected value of function. Thus, define the criterion as the maximization of the quantity

$$H_u = \overline{\ln\left[\frac{p(y|s)}{p(y)}\right]} = \int_{-\infty}^{\infty}\int p(y,s)\ln\left[\frac{p(y|s)}{p(y)}\right]dy\,ds \quad \ldots(6)$$

Equation (6) is recognized, of course, to be the mutual entropy between y and s or the useful entropy at y, a concept which has been widely used in other applications.

Arguments will now be developed to yield an alternative criterion to the maximization of useful entropy criterion.

Consider the linear system in Fig. 1.

Let the statistics of the signal and noise be $\bar{s} = \bar{n} = 0$, $\overline{sn} = 0$. Then, using capital letters to denote transforms ($E = \mathcal{F}(\varepsilon)$),

$$E = Y - S = (T-1)S + TN \quad \ldots\ldots(7)$$

$$\overline{ES} = (T-1)\overline{S^2} \quad \ldots\ldots(8)$$

Note that if $T = 1$,

$$E = TN \quad \ldots\ldots(9)$$

$$\overline{ES} = 0 \quad \ldots\ldots(10)$$

Setting $T = 1$ ensures that all of the input signal appears at the output y undistorted. The noise also appears, but there is no useful information content in the noise. Therefore, the maximum information about the desired signal 's' appears at the output 'y' in this case and consequently the minimum information about s appears in ε. Of course, this ideal situation cannot exist physically since T cannot be identically unity for all frequencies. Also, whereas the output information content at y (i.e. the mutual entropy between y and s) is independent of T, the error information content (i.e. the mutual entropy between ε and s) does depend on T. Thus, if the system is to be optimized by varying the parameters of T according to some optimization criterion, the maximization of output entropy cannot be used unless noise is also present on the output, whereas the minimization of error entropy can be applied regardless of the output noise conditions. The two criteria are clearly different. The output entropy maximization criterion is best when the error moments are desired small. The error entropy minimization is best when the signal is to be as undistorted as possible regardless of the noise. The error entropy is

$$H_u(\varepsilon) = \iint p(\varepsilon,s)\ln\frac{p(\varepsilon|s)}{p(\varepsilon)}\,d\varepsilon\,ds \quad \ldots\ldots(11)$$

Note that when $\overline{\varepsilon s} = 0$, $p(\varepsilon|s) = p(\varepsilon)$ and $H_u(\varepsilon) = 0$.

3. Entropy Calculations for Linear Systems

The ordinary communication system concerns itself with the properties of time functions and not with the spatial dimensions involved. Antenna system design, however, is concerned with both time and space functions. Hence the basic theory must be extended to include such functions. Consider the geometry of a typical system shown in Fig. 2.

In Fig. 2, the object coordinates in the object plane are x and y, in the aperture plane are u and v, and in the image plane are x' and y'. The object bounds are $-X \le x \le X$, $-Y \le y \le Y$, and similarly, the image bounds are $-X' \le x' \le X'$, $-Y' \le y' \le Y'$. These bounds are actually the largest physical dimension in each coordinate of the object and image, respec-

tively; that is, the bounds form a rectangle which circumscribes the object or image. The field strength at a point which lies within the circumscribing rectangle but outside the physical object or image is zero. The aperture is similarly bounded, i.e. $-U \leq u \leq U$, $-V \leq v \leq V$.

The polar coordinates of a point in the object plane are (R, θ, ϕ) and the polar coordinates of a point in the image plane are (R', θ', ϕ') where the origin of coordinates is 0 at the centre of the aperture. The optical axis is z.

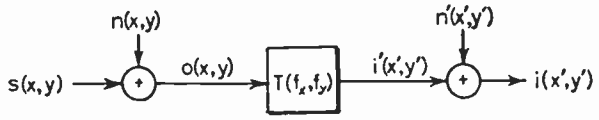


Fig. 3. Functional diagram of an antenna processing system.

Let us now consider the situation depicted in Fig. 3. Figure 3 represents an antenna system in which there is a signal and noise in the object which is filtered by the antenna spatial frequency filter and after which there is noise added in the image such that

$$o(x, y) = s(x, y) + n(x, y) \qquad \text{.....(12)}$$

$$i(x', y') = i'(x', y') + n'(x', y') \qquad \text{.....(13)}$$

Let us now determine what the probability distribution of the object field strength is. For convenience, consider one dimension only. An extended target may be thought of as a surface of elementary independent scatterers. This assumption is valid in a significant class of radar targets. It is most appropriate to discontinuous targets as found in applications such as a ground mapping. Returns which are approximated well by such a target model include ground and sea return, rain and chaff return, etc. Complex targets such as aircraft and missiles may also be approximated by such a model.[6] Although such targets in general will produce both diffuse and specular reflections, the specular reflections will usually be of finite time duration, unless the target is stationary. The approximation of purely diffuse reflection over the bounds of the target is probably most applicable when the target is in a state of motion in where there is significant vibration, giving rise to a scintillating signal.[7] Each scatterer reradiates the incident (temporally coherent) radiation with a change in amplitude and a shift in phase. Since only the scalar spatial problem is being considered, the polarization is assumed plane and unchanged by reflection. If the amplitude and phase for the ith scatterer are A_i and ϕ_i, respectively, the resultant signal at the antenna input is

$$\begin{aligned} o(x, t) &= \sum_{i=1}^{\infty} A_i(x) \cos [\omega t - \phi_i(x)] + n(x, t) \\ &= o_A(x) \cos \omega t - o_B(x) \sin \omega t \\ &= o(x) \cos [\omega t + \phi(x)] \end{aligned} \qquad \text{.....(14)}$$

where $o_A(x)$ and $o_B(x)$ are Gaussian by the central limit theorem and independent because the phasors $o_A \cos \omega t$ and $o_B \sin \omega t$ are orthogonal. The most reasonable *a priori* assumption is that $o_A(x)$ and $o_B(x)$ each have zero mean and the same variance. It therefore follows that $o(x, t)$ is Gaussian with mean zero and the same variance as $o_A(x)$ and $o_B(x)$, and that $o(x)$ is Rayleigh distributed and independent of $\phi(x)$ which is uniformly distributed over 2π. Note that

$$o_A(x) = o(x) \cos \phi(x)$$

$$o_B(x) = o(x) \sin \phi(x)$$

$o(x, t)$ consists of the sum of signal and noise. Let us assume that the signal and noise are statistically independent and have zero means.

The field strength in eqn. (14) is a narrow band Gaussian random process.[8] This representation follows from the basic assumption that the signal strength is made up of the return from many independent scatterers, each of which is illuminated by a signal which is sinusoidal in time. The randomness is introduced by the reflection from the scatterer and the summing of the independent returns. If the A_i and ϕ_i remain constant with time, the signal $s(x, t)$ is a sinusoid whose amplitude and phase are constant but *a priori* random because they are *a priori* unknown. The additive noise is also a narrow-band Gaussian process whose bandpass spectrum is narrow compared to ω. $o(x, t)$ is the sum of signal and noise and is therefore also a narrow-band Gaussian process.

It will be assumed that the object signal-plus-noise in eqn. (14) is demodulated coherently with a reference signal $2 \cos \omega t$. Coherent demodulation is here intended to mean the translation in frequency of the narrow bandpass spectrum of the signal in eqn. (14) from the high frequency ω_0 to 0 frequency. This is accomplished by multiplying $o(x, t)$ in eqn. (14) by $2 \cos \omega t$ and filtering out the high frequency terms. Let us therefore eliminate the temporal dependence of the object signal by coherently demodulating the function as suggested, so that the spatial object signal to be used for entropy calculations becomes

$$o(x) = s(x) + n(x) \qquad \text{.....(15)}$$

Since s and n are assumed Gaussian, all of the terms in eqn. (15) are Gaussian. It therefore follows that o in eqn. (14) is Gaussian and if the transfer function T in Fig. 3 is linear, then i' in eqn. (13) is Gaussian. We can further assume that the additive receiver noise or image noise n' is likewise Gaussian so that i in eqn. (13)

is also Gaussian. i and o in eqns. (12) and (13) are related by a convolution integral which is inconvenient for entropy calculations since the successive space samples are correlated. It is therefore more convenient to make use of frequency samples. Frequency samples will be denoted by capital letters and space samples by lower case letters. The total useful entropy (sometimes called trans-information or mutual information) is

$$\begin{aligned} H_{ut}(i) &= H(i_1, i_2, \ldots, i_M) - \\ &\quad - H(i_1, i_2, \ldots, i_M | s_1, s_2, \ldots, s_M) \\ &= H(I_1, I_2, \ldots, I_M) - \\ &\quad - H(I_1, I_2, \ldots, I_M | S_1, S_2, \ldots, S_M) \end{aligned} \quad (16)$$

Note that the total useful entropy is the same regardless of whether space samples or frequency samples are used. In eqn. (16), M represents the total number of samples. The general expression (16) can be much simplified if the successive samples are independent or if the process is stationary or both. The evaluation of the entropy for the case described by Fig. 3 will be evaluated first for the most general situation where none of these simplifying conditions hold, and then the effect of the simplifying assumptions will be shown. In order to evaluate the entropy (16), the relationship between I and S, N, N' must be derived. Thus:

$$\begin{aligned} I(f_x, f_y) &= I'(f_x, f_y) + N'(f_x, f_y) \\ &= T(f_x, f_y)[S(f_x, f_y) + N(f_x, f_y)] + N'(f_x, f_y) \end{aligned} \quad (17)$$

In order to evaluate the total entropy, the joint distribution $p(I_1, I_2, \ldots, I_M, S_1, S_2, \ldots, S_M)$ must be found. This distribution may be found by functional transformation from the known joint distribution of the input variables. Thus, the total useful entropy at the output of the system in Fig. 3 is

$$\begin{aligned} H_{ut} &= -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(I) \ln p(I) \prod_{i=1}^{M} \mathrm{d}I_i \\ &\quad + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(I, S) \ln p(I|S) \prod_{i=1}^{M} \mathrm{d}I_i \, \mathrm{d}S_i \end{aligned}$$

$$H_{ut} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(I, S) \ln \left[\frac{p(I|S)}{p(I)} \right] \prod_{i=1}^{M} \mathrm{d}I_i \, \mathrm{d}S_i \quad \ldots\ldots(18)$$

Let us now find the total image useful entropy for particular distributions assumed for the signal and noise inputs. For example, assume that the inputs (namely, $s(x)$, $n(x)$ and $n'(x)$) are each Gaussian with 0 mean and a known spectral density defined over all spatial frequencies. These specifications are sufficient to completely specify the probability distribution, since the mean and variance give a complete statistical description of a one dimensional Gaussian variable. For a multi-dimensional Gaussian distribution, the second-order moment matrix must be specified, that is, all of the correlation coefficients between the various sample variables as well as their variances. From the Appendix, eqn. (52), the total useful entropy for a multidimensional Gaussian distribution is

$$H_{ut} = \tfrac{1}{2} \ln \left(\frac{\Lambda_I}{\Lambda_{I|S}} \right) \quad \ldots\ldots(19)$$

where

$$\Lambda_{I|S} = \det [\lambda_{ij}(I|S)] \quad \ldots\ldots(20)$$

$$\Lambda_I = \det [\lambda_{ij}(I)] \quad \ldots\ldots(21)$$

$$\lambda_{ij}(I|S) = T_i T_j^* \rho_{N_i N_j} \sigma_{N_i} \sigma_{N_j} + \rho_{N_{i'} N_{j'}} \sigma_{N_{i'}} \sigma_{N_{j'}} \quad \ldots\ldots(22)$$

$$\lambda_{ij}(I) = T_i T_j^* \rho_{S_i S_j} \sigma_{S_i} \sigma_{S_j} + \lambda_{ij}(I|S) \quad \ldots\ldots(23)$$

The subscripts i and j refer to frequency sampling instants, T_i is the filter transfer function at the ith frequency, $\rho_{A_i A_j}$ is the correlation coefficient between A_i and A_j, where A_i is a signal, image noise, or object noise frequency sample at the ith frequency and $\sigma_{A_i}^2$ is the variance of this frequency sample. If the successive samples are uncorrelated, although nonstationary in frequency, a very great simplification occurs. In this case, the correlation coefficients become

$$\rho_{S_i S_j} = \rho_{N_i N_j} = \rho_{N_{i'} N_{j'}} = \delta_{ij} \quad \ldots\ldots(24)$$

where δ_{ij} is the Kronecker δ. Equations (22) and (23) then reduce to

$$\lambda_{ii}(I|S) = |T_i|^2 \sigma_{N_i}^2 + \sigma_{N_{i'}}^2 \quad \ldots\ldots(25)$$

$$\lambda_{ii}(I) = |T_i|^2 \sigma_{S_i}^2 + \lambda_{ii}(I|S) \quad \ldots\ldots(26)$$

Notice that only the diagonal terms in the appropriate moment matrices remain different from 0 (they are specified by eqns. (25) and (26)) so that eqn. (19) becomes

$$\begin{aligned} H_{ut} &= \tfrac{1}{2} \ln \left(\frac{\prod_{i=1}^{m} \lambda_{ii}(I)}{\prod_{i=1}^{n} \lambda_{ii}(I|S)} \right) = \tfrac{1}{2} \sum_{i=1}^{m} \ln \left(\frac{\lambda_{ii}(I)}{\lambda_{ii}(I|S)} \right) \\ &= \tfrac{1}{2} \sum_{i=1}^{m} \ln \left(\frac{|T_i|^2 (\sigma_{S_i}^2 + \sigma_{N_i}^2) + \sigma_{N_{i'}}^2}{|T_i|^2 \sigma_{N_i}^2 + \sigma_{N_{i'}}^2} \right) \\ &= \tfrac{1}{2} \sum_{i=1}^{m} \ln \left[1 + \frac{\sigma_{S_i}^2}{\sigma_{N_i}^2 + \frac{\sigma_{N_{i'}}^2}{|T_i|^2}} \right] \end{aligned} \quad \ldots\ldots(27)$$

Note that the signal spectral density is given by

$$G_s(f_x, f_y) = \frac{\overline{|S(f_x, f_y)|^2}}{4XY} = \frac{\sigma_s^2}{4XY} \quad \ldots\ldots(28)$$

G_s is the spatial frequency spectral density of the object signal strength. The corresponding spectral densities for the input and output noise are, respectively,

$$G_n(f_x, f_y) = \frac{\overline{|N(f_x, f_y)|^2}}{4XY} = \frac{\sigma_N^2}{4XY} \quad \ldots\ldots(29)$$

$$G'_n(f_x, f_y) = \frac{\overline{|N'(f_x, f_y)|^2}}{4XY} = \frac{\sigma^2_{N'}}{4XY} \quad \text{.....(30)}$$

In eqns. (28) and (29) it has been assumed that the desired or signal object dimensions ($\pm X$, $\pm Y$) are the same as the object and image noise dimensions.

Substituting eqns. (28)–(30) in eqn. (27) yields

$$H_{ut} = \tfrac{1}{2}\sum_{i=1}^{M} \ln\left[1 + \frac{G_s(f_i)}{G_n(f_i) + \dfrac{G'_n(f_i)}{|T(j\omega_i)|^2}}\right] \quad \text{...(31)}$$

where $f = (f_x, f_y)$.

4. Antenna Optimization

Since the antenna spatial frequency filter is a low-pass passive filter,

$$|T(j\omega)| \leq 1 \quad \text{.....(32)}$$

According to the first criterion the output total useful entropy and useful entropy per sample are to be maximized. From eqn. (31) it is clear that the output total useful entropy is maximized if the bounds $B_u B_v$ on the spatial frequency filter are as large as possible and if the equality in eqn. (32) holds over the frequency bounds of the aperture. In other words, in the absence of any other constraints, the total useful entropy in the output is maximized if

$$T(j\omega) = [U(f_x + B_u) - U(f_x - B_u)] \times \times [U(f_y + B_v) - U(f_y - B_v)] \quad (33)$$

where $U(x)$ is the unit step function.

Thus, the total output useful entropy is maximized if the aperture illumination is uniform and if the aperture dimensions are as large as possible. Since both the object signal spectrum and object noise spectrum are broad compared to the aperture filter width, maximizing the entropy per sample requires that the equality in eqn. (32) hold, but puts no requirement on the aperture bounds.

It is now of interest to apply the second criterion. In this case, the system error is related to the three inputs in the following way:

$$\begin{aligned} E(f_x, f_y) &= S(f_x, f_y) - I(f_x, f_y) \\ &= [1 - T(f_x, f_y)]S(f_x, f_y) - \\ &\quad - T(f_x, f_y)N(f_x, f_y) - N'(f_x, f_y) \end{aligned} \quad (34)$$

Again, if all of the input distributions are Gaussian, E is Gaussianly distributed, then by analogy with eqn. (19), the total useful entropy is

$$H_{ut} = \tfrac{1}{2}\ln\left(\frac{\Lambda_E}{\Lambda_{E|S}}\right) \quad \text{.....(35)}$$

where

$$\bar{E}_i|S_i = (1 - T_i)S_i \quad \text{.....(36)}$$

$$E_i - \bar{E}_i|S_i = -T_i N_i - N'_i \quad \text{.....(37)}$$

$$\begin{aligned} \Lambda_{E|S} &= [\lambda_{ij}(E|S)] = \overline{[(E_i - \bar{E}_i|S_i)(E^*_j - \bar{E}^*_j|S_j)]} \\ &= \overline{[(T_i N_i + N'_i)(T^*_j N_j + N'_j)]} \\ &= [T_i T^*_j \rho_{N_i N_j} \sigma_{N_i} \sigma_{N_j} + \rho_{N_i N'_j} \sigma_{N_i} \sigma_{N_j}] \end{aligned} \quad (38)$$

$$\begin{aligned} \Lambda_E &= \overline{[(E_i - \bar{E}_i)(E_j - \bar{E}_j)]} = [\lambda_{ij}(E)] \\ &= [(1 - T_i)(1 - T_j)^* \overline{S_i S^*_j} + \lambda_{ij}(E|S)] \\ &= [(1 - T_i)(1 - T_j)^* \rho_{S_i S_j} \sigma_{S_i} \sigma_{S_j} + \lambda_{ij}(E|S)] \end{aligned} \quad \text{.....(39)}$$

Exactly the same arguments which applied in finding the output useful entropy can be used in finding the useful error entropy. When the successive frequency samples are independent but not frequency stationary, by analogy with eqn. (27), the total useful entropy becomes

$$\begin{aligned} H_{ut} &= \tfrac{1}{2}\sum_{i=1}^{M} \ln\left(\frac{\lambda_{ii}(E)}{\lambda_{ii}(E|S)}\right) \\ &= \tfrac{1}{2}\sum_{i=1}^{M} \ln\left(1 + \frac{|1 - T_i|^2 \sigma^2_{S_i}}{|T_i|^2 \sigma^2_{N_i} + \sigma^2_{N_{i'}}}\right) \end{aligned} \quad \text{.....(40)}$$

Employing eqns. (28)–(30), eqn. (40) can be written

$$H_{ut}(\varepsilon) = \tfrac{1}{2}\sum_{i=1}^{M} \ln\left[1 + \frac{|1 - T_i|^2 G_s(f_i)}{|T_i|^2 G_n(f_i) + G'_n(f_i)}\right] \quad \text{...(41)}$$

In the absence of any constraints except that of eqn. (32), $H_{ut}(\varepsilon)$ is minimized by setting $|T(j\omega)| = 1$ over the aperture, where the aperture bounds are as large as possible. This same conclusion was reached by the output information maximization criterion.

When $T(f)$ assumes a simple form, as in a linear, one-way antenna, the general conclusions are that $T(f)$ should be as large as possible and as uniform as possible over all frequencies. Since for a passive antenna, the transfer function $|T| \leq 1$, the antenna is optimized with respect to both criteria by setting $|T| = 1$, i.e. specifying uniform illumination. This conclusion can be reached only for this very simple situation; for the more realistic case of a two-way pattern, one cannot simultaneously satisfy uniformity

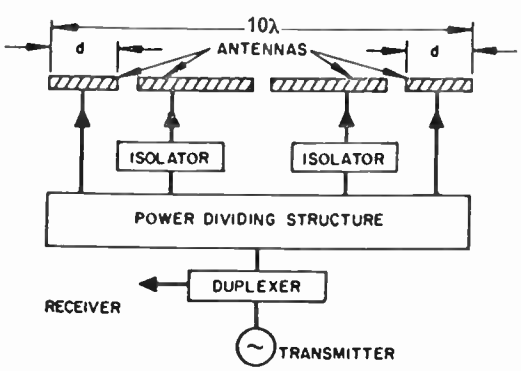


Fig. 4. Non-reciprocal array.

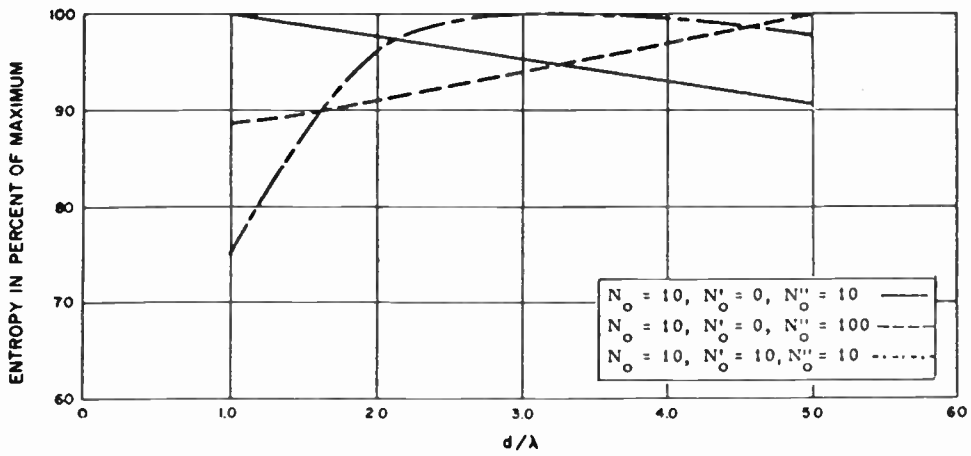


Fig. 5. Output entropy for $S_o = 1$ and the noise conditions indicated in the illustration.

and maximality. If, moreover, one permits a non-reciprocal system (in which the transmitting and receiving antennas differ) then the optimum antenna parameters will be highly dependent on the signal and noise characteristics of the object and image.

For a reciprocal, two-way, one-dimensional linear antenna, the requirement is that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} g^2(x)\,e^{j\omega x}\,dx = U(f+B)-U(f-B) \quad ...(42)$$

or, that

$$g(x) = \text{const}\sqrt{\frac{\sin \pi Bx}{\pi Bx}} \quad .....(43)$$

where $g(x)$ is the pattern for an aperture of spatial frequency range $2B$. The function in eqn. (43) is quite difficult to approximate by ordinary aperture synthesis. The required aperture distribution will be concave upward. An exact realization can be achieved by using a non-reciprocal system.

To illustrate the concepts and techniques developed in the paper, an example of optimizing a two-way, non-reciprocal, antenna was carried out. A diagram of the antenna system appears in Fig. 4. On transmit, the 10λ array is uniformly illuminated, while on receive, an interferometer array of variable element length d was allowed to vary, and the optimum d determined according to the two criteria. The output criterion shows the optimum d to be a function of the signal and noise conditions. For example, let the spectra be flat, so that $G_n(f) = N_o$, $G_n'(f) = N_o' + N_o''$ where N_o' is amplifier noise and N_o'' is processing system noise, and $G_s(f) = S_o = 1$. Curves for representative values of N_o, N_o' and N_o'' appear in Fig. 5. Note that the maximum entropy is achieved for different d depending on the noise environment.

5. Conclusions

In this paper a fundamental theory has been developed for the evaluation and optimization of systems carrying spatial as well as temporal information. A significant development which was essential to the above theory was the extension of temporal information theory to the four-dimensional space-time domain. The principal optimization criteria developed are maximization of information content and rate and minimization of information content and rate in the error (deviation between desired and actual output). These criteria are compared to other more conventional measures of quality such as minimization of r.m.s. error and maximization of output signal/noise ratio. The proposed criteria are shown to be superior in several respects to the conventional ones.

The initial conclusions from the above study are that the most important characteristics to be prescribed for the spatial frequency transfer function of an antenna are uniformity (with frequency) and minimum attention. Both of these requirements can be simultaneously satisfied in a one-way pattern by prescribing uniform illumination. In the case of two-way patterns a certain trade-off between the two requirements must be accepted. In the case of reciprocal antennas the uniformity requirement cannot be met. It is possible, however, to achieve it with a non-reciprocal array, although at the expense of increased attenuation. The optimum antenna configurations were obtained for various signal and noise conditions.

6. Acknowledgment

The research reported in this paper was sponsored by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract AF 19(628)–333.

7. References and Bibliography

1. P. B. Fellgett and E. H. Linfoot, "On the assessment of optical images", *Phil. Trans. Roy. Soc. London*, Series A, **247**, No. 931, pp. 369–407, 1954.
2. D. Gabor, "Light and information", in "Progress in Optics", Vol. 1, edited by E. Wolf. (North Holland Publishing Co., 1961.)
3. H. Walter, "On basic analogies and principal differences between optical and electronic information", in "Progress in Optics", Vol. 1, edited by E. Wolf. (North Holland Publishing Co., 1961.)
4. G. O. Young and A. Ksienski, "Signal and data-processing antennas", *I.R.E. Trans. on Military Electronics*, **MIL-5**, No. 2, pp. 94–102, April 1961.
5. E. L. O'Neill, R. Barakat and D. E. Lev, "Communication Theory and Scalar Diffraction Theory", Itek Corp., Final Report, No. P-168, April 1960.
6. R. H. Delano, "A theory of target glint or angular scintillation in radar tracking", *Proc. Inst. Radio Engrs*, **41**, pp. 1178–84, December 1953.
7. Robert B. Muchmore, "Aircraft scintillation spectra", *I.R.E. Trans. on Antennas and Propagation*, **AP-8**, No. 2, pp. 201–12, March 1960.
8. W. B. Davenport and W. L. Root, "Introduction to Random Signals and Noise", p. 158. (McGraw-Hill, New York, 1958.)

8. Appendix

The Useful Entropy for a Multi-dimensional Gaussian Distribution

The n-dimensional Gaussian distribution is given by

$$p(Z_1, Z_2, \ldots, Z_n) = \frac{1}{(2\pi)^{n/2}\Lambda^{1/2}} \times \times \exp\left\{-\frac{1}{2\Lambda}\sum_{i,j=1}^{n}\Lambda_{ji}(Z_i-\bar{Z}_i)(Z_j-\bar{Z}_j)\right\} \quad (44)$$

where

$$\Lambda = (\lambda_{ij}) = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{n1} & \lambda_{n2} & \ldots & \lambda_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & & \ldots & \rho_{2n}\sigma_2\sigma_n \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \ldots & \sigma_n^2 \end{pmatrix} \quad \ldots\ldots(45)$$

and

$$\lambda_{ij} = \overline{(Z_i-\bar{Z}_i)(Z_j-\bar{Z}_j)} = \rho_{ij}\sigma_i\sigma_j \quad \ldots\ldots(46)$$

$$\Lambda = \det\Lambda = \begin{vmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{n1} & \lambda_{n2} & \ldots & \lambda_{nn} \end{vmatrix} \quad \ldots\ldots(47)$$

$$\Lambda_{ij} = \mathrm{cof}(\lambda_{ij}) \quad \ldots\ldots(48)$$

Therefore

$$\begin{aligned} H(Z) &= -\int_{R_Z}\prod_{i=1}^{n} dZ\, p(Z)\ln p(Z) = -\overline{\ln p(Z)} \\ &= \ln\left[(2\pi)^{n/2}\Lambda^{1/2}\right] + \frac{1}{2\Lambda}\sum_{i,j=1}^{n}\Lambda_{ji}\overline{(Z_i-\bar{Z}_i)(Z_j-\bar{Z}_j)} \\ &= \ln\left[(2\pi)^{n/2}\Lambda^{1/2}\right] + \frac{1}{2\Lambda}\sum_{i,j=1}^{n}\Lambda_{ji}\lambda_{ij} \\ &= \ln\left[(2\pi)^{n/2}\Lambda^{1/2}\right] + \tfrac{1}{2}\sum_{j=1}^{n}U_{jj} \\ &= \ln\left[(2\pi)^{n/2}\Lambda^{1/2}\right] + \frac{n}{2} \\ &= \tfrac{1}{2}\ln\left[(2\pi e)^n\Lambda\right] \quad \ldots\ldots(49) \end{aligned}$$

The following facts were used in the derivation of eqn. (49):

$$\sum_{i=1}^{n}\frac{\lambda_{ji}\lambda_{ij}}{\Lambda} = U_{jj} \quad \ldots\ldots(50)$$

where U_{ij} is an element of the unit matrix

$$U = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 \end{pmatrix} \quad \ldots\ldots(51)$$

Substituting eqn. (49) in eqn. (16) yields

$$H_{ut} = \tfrac{1}{2}\ln\left(\frac{\Lambda_I}{\Lambda_{I|S}}\right) \quad \ldots\ldots(52)$$

where

$$\left.\begin{aligned} \Lambda_{I|S} &= |\lambda_{ij}(I|S)| \\ \Lambda_I &= |\lambda_{ij}(I)| \end{aligned}\right\} \quad \ldots\ldots(53)$$

and

$$\lambda_{ij}(I|S) = \overline{[I_i-\bar{I}_i|S_i][I_j^*-\bar{I}_j^*|S_j]|S_i, S_j} \quad \ldots(54)$$

$$\lambda_{ij}(I) = \overline{[I_i-\bar{I}_i][I_j^*-\bar{I}_j^*]} \quad \ldots\ldots(55)$$

But, from eqn. (17),

$$\bar{I}_i|S_i = T_iS_i \quad \ldots\ldots(56)$$

Hence

$$\begin{aligned} \lambda_{ij}(I|S) &= \overline{(T_iN_i+N_i')(T_j^*N_j^*+N_j'^*)} \\ &= T_iT_j^*\overline{N_iN_j^*} + \overline{N_i'N_j'^*} \\ &= T_iT_j^*\rho_{N_iN_j}\sigma_{N_i}\sigma_{N_j} + \rho_{N_i'N_j'}\sigma_{N_i'}\sigma_{N_j'} \quad \ldots(57) \end{aligned}$$

From eqn. (17), $\bar{I}_i = 0$. Therefore

$$\begin{aligned} \lambda_{ij}(I) = \overline{I_iI_j^*} &= \overline{[T_i(S_i+N_i)+N_i'][T_j^*(S_j^*+N_j^*)+N_j'^*]} \\ &= T_iT_j^*[\rho_{S_iS_j}\sigma_{S_i}\sigma_{S_j} + \rho_{N_iN_j}\sigma_{N_i}\sigma_{N_j}] + \\ &\quad + \rho_{N_i'N_j'}\sigma_{N_i'}\sigma_{N_j'} \quad \ldots(58) \end{aligned}$$

Manuscript first received by the Institution on 31st March 1964, and in final form on 30th October 1964. (Paper No. 962.)

DISCUSSION

Under the chairmanship of Mr. R. N. Lord

Mr. J. Croney: Can Dr. Young give a physical picture of why in a one-way system one optimizes information with a uniform distribution but a two-way system requires to be uniform one way and non-uniform the other. This seems to be not in accord with the reciprocity theorem and the physical picture does not emerge from the mathematics.

Dr. G. O. Young (*in reply*): Maximum information is achieved at the output of an antenna when the object field is distorted as little as possible by the antenna. In a two-way system, the object field is already modified by the antenna pattern of the illuminating energy, so that after reception, the object field has been effectively multiplied by the product of the transmitting antenna pattern and the receiving antenna pattern. This product is equivalent to a convolution in the spatial frequency domain. In a one-way system, the object field is multiplied by the receiving antenna pattern only. Therefore, in the one-way case, the aperture shading should be uniform whereas in the two-way case the transmitting antenna illumination convolved with the receiving antenna illumination should be uniform. It is not necessary that either of these be uniform; however, if one of them is uniform, the other must necessarily be non-uniform in order that their convolution be uniform.

These conclusions actually have nothing to do with the classical reciprocity theorem of electromagnetic theory. This theorem says essentially that the transmitting and receiving patterns of a linear antenna are the same. The antenna which is used in the paper is non-reciprocal in the sense that isolators are used so that one pattern is used on transmit and another on receive. It is more instructive, however, to consider this arrangement as two separate antennas, one used for transmit and the other for receive. The reciprocity theorem applies to each of these antennas separately.

Professor M. Ryle: I believe that the system must be reciprocal: surely one can obtain the desired aperture distribution on reception by the use of loss-free transformers instead of attenuators?

In the case of radio astronomy mapping, one would not want to use a uniform aperture illumination; the best illumination will depend on the relative number of sources at different intensity levels, but a function leading to much smaller side-lobes will be needed.

Dr. Young (*in reply*): Although loss-free transformers could be used instead of attenuators on receive, I do not believe that the desired amplitude shading can be achieved without attenuation. The reason for this is the change in impedance level introduced by step-up or step-down transformers. Even though loss-free transformers are used, the source impedance on the input to the transformer (times the square of the turns ratio) will be transformed to the transformer output. When the various signals are combined across a common load impedance, the maximum signal power transfer and hence the maximum signal/noise ratio is achieved when there is a match between the source and load impedances. Regardless of whether the various voltage levels are achieved by voltage dividers or transformers, that portion of the power which is not desired across the summing impedance must be dissipated.

The conclusion that uniform aperture illumination should be used to achieve maximum information content on the output of the antenna system was based on the assumption that the signal power was uniformly distributed over the sky. If the signal power is not so distributed, as is usually the case in radio astronomy, the method proposed in the paper can be used to determine what the optimum shading should be, but this shading will not necessarily be uniform.

Mr. P. M. Woodward: A questioner seemed puzzled by the optimal property of a uniformly-illuminated aperture, because it has high side-lobes. In any problem of Fourier analysis with a limited 'aperture' or a limited number of Fourier terms, the unaltered low-order terms give the best least-squares fit. Bad though the side-lobes may be, any other than uniform illumination would give an even worse fit.

Dr. Young (*in reply*): I agree with Mr. Woodward's comment. It is interesting that in the example used in the paper, uniform illumination not only gives the best least squares fit, but also yields the highest information content.

Dr. R. Benjamin: For optimum received signal/noise ratio from an array of aerial elements, each contribution ought to be (losslessly) weighted in proportion to its coherent signal voltage and in inverse proportion to its non-coherent noise power. Hence, by reciprocity, optimum target illumination requires a uniform aerial aperture illumination.

On reception, the *angular* signal information is proportional to the spacing of the aerial element from the centre of the array. Hence, angular information is optimized by a V-shaped voltage taper, i.e. a V-shaped power taper. However, target-existence information, against a background of thermal noise coupling to the array elements, would require an untapered distribution. Hence, for maximum information, the aerial distribution, on reception, should be adaptively matched to the signal/noise ratio pertaining. This can be done by transformers at the array elements or by attenuation following independent amplifiers for the array elements. Independent recording of the signals from the individual array elements would permit retrospective optimization in the processing system.

Thus, roughly speaking, when there is little noise, the full aperture is available as an interferometer. When the signal/noise ratio is poor, more of the available aperture must be used to intercept the incoming waveform. Hence, the interferometer elements must be extended inwards, and so their effective mean spacing is reduced.

Dr. Young (*in reply*): Dr. Benjamin has presented an excellent physical explanation for the conclusions reached

in the paper, assuming a uniformly illuminated antenna on transmit. It is interesting that an interferometer is the limiting case of a 'V-shaped taper'.

Mr. K. F. Molz: The AN-FPS 85 already uses uniform illumination on transmit to get higher power on target and space taper on receive to obtain low overall side-lobe performance.†

Mr. C. van Schooneveld: Is it possible to extend the computations to the case of a background noise which is spatially non-stationary in two-dimensional object space?

Dr. Young (*in reply*): Yes, it is definitely possible to carry out the information content and rate computations for the case of spatially non-stationary objects and/or background noise and to determine the optimum antenna under such circumstances. For simplicity, I chose an example where the background noise was spatially stationary. If the process under consideration is not stationary, the information content at each spatial sample is simply summed to give the total information content, providing the successive samples are statistically independent. If the process is spatially stationary as well as consisting of statistically independent samples, the contribution to the output information content from each sample is the same. The total content is then found by multiplying the information content arising from one sample by the number of samples. The most difficult case occurs when the process is such that the successive spatial samples are not statistically independent. In this case, the total information content is found from the joint probability distribution of all of the samples.

Mr. P. Bradsell: Is there a simple physical interpretation of the result obtained for the two-way non-reciprocal array? The analysis indicates that the whole aperture should be used for transmission, and part of the aperture for reception. Would, for instance, a radar equipment give improved performance if operated in this way?

Dr. Young (*in reply*): This question is answered in part in the answer to Mr. Croney's question. Dr. Benjamin's remarks are also very appropriate here. The reason that part of the aperture is used for reception in the non-reciprocal two-way array is that this is required in order to give an overall uniform two-way transfer function. The disadvantage of using only part of the aperture on reception is that the signal strength and consequently the signal/receiver-noise ratio is degraded as a consequence. It is this effect which decides what portion of the aperture shall be used in order to yield the maximum output information content. Several signal and noise conditions were considered in the paper and some typical curves plotted. Thus, the use of such an array would give improved radar performance under certain signal and noise conditions.

Dr. K. Milne: It is difficult to reconcile the concept that a uniformly illuminated array provides maximum information with the fact that it has high side-lobes which must represent incorrect information. Could the author resolve the paradox?

The enhancement process in Dr. Schell's paper† processes the data to provide a response which has very low side-lobes. Could Dr. Young give some numerical indication of the loss of information arising when a uniformly illuminated aperture is replaced by one with, say, a triangular illumination?

Dr. Young (*in reply*): Dr. Milne's first question is answered in the response to one of Professor Ryle's questions. Briefly, the uniformly illuminated array provides maximum information for a target whose power is uniformly distributed over the sky. It might be argued that this is not a physically reasonable source. It is, however, a reasonable target model for example for point targets whose *a priori* location in the sky is unknown and equally likely to occur anywhere. Thus, if the antenna is to maximize the information content over a relatively long period of time, although a single point target will appear at only one place in the sky at a given instant of time, if the totality of targets appearing over an interval of time is considered as the source, then a uniform distribution of signal power over the sky is a reasonable assumption. If such a source distribution is considered, the high side-lobes do not represent incorrect information, and indeed yield more information than a system with lower side-lobes. If *a priori* target information is available, the same techniques employed in the paper could well lead to different conclusions than those reached in the example.

Since a uniformly illuminated aperture is equivalent to a spatial frequency transfer function whose attenuation characteristic is constant over the aperture and since a triangular illumination corresponds to a triangular filter, there will be a loss of information arising from the use of a triangular filter in place of a uniform filter. Indeed, this exact problem has been worked out by Claude Shannon‡ who shows that the loss in entropy is $1/e^2$, or 8·68 dB loss in 'entropy power' (i.e. information content) as compared to the corresponding uniform filter attenuation characteristic.

Dr. A. A. Ksienski (*communicated*): I should like to comment on the controversy which seems to have arisen regarding the conclusion that uniform illumination with its ensuing high side-lobes is the optimum distribution. The controversy stems partly from the different functions that different people expect the antenna to perform. If one expects the antenna to be just an information-gathering device which transforms the incoming waves (electromagnetic, acoustic or other) into signals of more desirable form for processing, then obviously the optimum antenna is the one which captures the maximum amount of information and transforms it with the least amount of distortion. It is in this sense that a uniform illumination is optimum. If, on the other hand one expects the antenna to be a processor as well as a gatherer of information,

† K. F. Molz, "Phased array radar systems", *The Radio and Electronic Engineer*, **27**, No. 5, pp. 339–48, November 1964.

† A. C. Schell, "Enhancing the angular resolution of incoherent systems", *The Radio and Electronic Engineer*, **29**, No. 1, pp. 21–6, January 1965.

‡ Claude E. Shannon and Warren Weaver, "The Mathematical Theory of Communication", page 61. (University of Illinois Press, 1949.)

certain trade-offs become necessary. Thus, if one expects the output of the antenna to be directly conveyed to a display and that display to be the explicit description of the object field, the criterion of performance will be the minimum deviation of the visual display, or the image, from the object. This is similar to the very familiar requirements of obtaining the best approximation to a function (or a pattern synthesis) by means of a finite trigonometric sum which represents the array. The best fit in the present case will quite likely not be in the mean square sense due to the Gibbs' phenomenon (i.e. high side-lobes), but the approximation which will yield the least absolute deviation i.e. the Chebyshev type approximation. Now, this does not mean that the Fourier approximation or uniform illumination does not carry the desired information—it most certainly does since the presently optimum illumination is obtained by attenuating certain of the array elements (i.e. tapering), or deliberately destroying part of the implicit information in the signal for the convenience of explicit representation on a display.

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

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

January 1965	GBR 16kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	January 1965	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	— 149·6	— 150·5	— 1	17	— 148·0	— 149·1	— 6
2	— 150·3	— 148·7	0	18	— 147·9	— 150·6	— 7
3	— 149·6	— 148·9	0	19	— 151·5	— 151·3	— 7
4	— 149·8	— 149·6	— 7	20	— 150·3	— 149·6	— 5
5	— 150·4	— 150·1	— 9	21	— 149·6	— 150·5	— 6
6	— 149·6	—	— 9	22	— 150·0	— 150·7	— 5
7	— 149·4	— 150·1	— 8	23	— 150·0	— 149·8	— 5
8	—	— 149·7	— 9	24	— 150·4	— 150·9	— 5
9	— 150·5	— 150 0	— 7	25	— 151·1	— 152 6	— 4
10	— 149·7	— 149·0	— 8	26	— 151·2	— 151·9	— 4
11	— 149·1	— 151·3	— 9	27	— 150·8	— 150·3	— 3
12	— 150·1	— 150·3	— 7	28	— 150·5	— 149·4	— 4
13	— 150·6	— 149·9	— 7	29	— 150·4	— 150·1	— 4
14	— 149·8	— 151·6	— 7	30	— 149·8	— 150·1	— 4
15	— 149·9	— 150·9	— 7	31	— 149·7	— 149·6	— 3
16	— 149·8	— 148·3	— 6				

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium F,m (4,0)–F,m (3,0) transition at zero field.

Notes: (1) The phase of the GBR and MFS time signals was retarded by 100 milliseconds at 00 00 U.T. on 1st January 1965 and will be retarded by a further 100 milliseconds at 00 00 U.T. on 1st March 1965.

(2) The frequency offset for 1965 will be -150×10^{-10} c/s.

I.E.R.E. GRADUATESHIP EXAMINATION, NOVEMBER 1964

PASS LISTS

The following candidates who sat the November 1964 examination at centres outside Great Britain and Ireland succeeded in the sections indicated. The examination, which was conducted at 66 centres throughout the world, attracted entries from 459 candidates. Of these 134 sat the examination at centres in Great Britain and Ireland and 169 sat the examination at centres overseas. The names of successful candidates resident in Great Britain and Ireland are published in the January/February issue of the *Proceedings* of the I.E.R.E.

	Candidates appearing	*Pass*	*Fail*	*Refer*
Section A				
Great Britain	67	28	30	9
Overseas	93	23	62	8
Section B				
Great Britain	67	22	28	17
Overseas	76	10	48	18

OVERSEAS

The following candidates have now completed the Graduateship Examination and thus qualify for transfer or election to Graduate or a higher grade of membership.

DESAI, S. B. (S), *Bangalore.*
HACKING, J. B. (S), *Bermuda.*
KUSHWAHA, R. C. (S), *Lucknow.*
LEE, H. W. (S), *Hong Kong.*
OLIVER, T. J. (S), *Suva.*
RAMACHANDRA, K. (S), *Bangalore.*
SCHER, J. (S), *Tel-Aviv.*
SHARPE, M. W. V., *Hong Kong.*
SINGH, K. (S), *Bangalore.*
TAN, J. K., *Kuala Lumpur.*

The following candidates have now satisfied the requirements of Section A of the Graduateship Examination.

ABU, W. (S), *Lagos.*
AGALA, M. (S), *Lagos.*
AGBASHE, E. O. (S), *Lagos.*
AGBOLA, T. O. (S), *Lagos.*
AQL, N. H. M., *Cairo.*
CHEN, S. P., *Kuala Lumpur.*
COOPER, R. A. (S), *R.A.F., Seletar.*
DEEKSHITULU, R. S. V. Y. (S), *Hyderabad.*
FURDAL, B. O. (S), *Bergen.*
GUHA BARMAN, J. K. (S), *Calcutta.*
GURUDATH, D. N. (S), *Bangalore.*
KHAN, M. Y. (S), *Lahore.*
MUNDAY, J. D., *Barbados.*
NAGAPPA, N. G. (S), *Bangalore.*
OJI, E. D. (S), *Lagos.*
PANDEY, A. K. (S), *Dehra Dun.*
PAUL, R. N., *Agra.*
PRASAD, V. S., *Hyderabad.*
RADHAKRISHNAN UNNI, A. K., *Madras.*
RAMACHANDRA IYER, H., *Calcutta.*
RAMACHANDRA MENON, M., *Bombay.*
SHUJA-UD-DIN (S), *Rawalpindi.*
YAP, C. S., *Singapore.*

The following name was omitted from the list of candidates who qualified at the May 1964 examination, as published in *The Radio and Electronic Engineer*, page 292, October, 1964:

MULTANI, S. G. (S), *Bombay.*

The question papers set in Section A of the May 1964 Graduateship Examination were published in the October/December issue of the *Proceedings* of the I.E.R.E., together with answers to numerical questions and examiners' comments. Parts 3 and 4 of Section B and Part 5 of Section B will be published in subsequent issues of *Proceedings.*

(S) denotes a Registered Student.

U.D.C. 621.372.852.32

Coupled-wave Description of the Absorption-type Ferrite Modulator

By

J. HELSZAJN, M.S.E.E.
(*Graduate*)†

Summary: Coupled wave theory is applied to the absorption type ferrite modulator. The importance of the quantities $(\alpha_x-\alpha_y)/K$ and $(\beta_x-\beta_y)/K$ in determining the attenuation of the incident polarization are thereby demonstrated. For maximum attenuation we require $(\alpha_x-\alpha_y)/K = -2$ and $(\beta_x-\beta_y)/K = 0$. Some experimental data confirming the effect of the quantity $(\alpha_x-\alpha_y)/K$ on the loss of the incident polarization are given.

1. Introduction

It is the purpose of this paper to evaluate the absorption-type ferrite attenuator[1, 2] in terms of coupled-wave theory. This ferrite attenuator is characterized by large reciprocal attenuation over a broad band of frequencies and compares favourably with the Faraday rotator type which is limited to about 20 dB because of ellipticity. The attenuator consists of two ferrite slabs centrally located in a standard rectangular waveguide between which is sandwiched a thin resistance card perpendicular to the input r.f. electric field. The ferrite is biased by a longitudinal magnetic field applied through a solenoid wound around the waveguide. In the absence of a d.c. field, the r.f. energy is transmitted without attenuation. In the magnetized state, the r.f. wave is attenuated. Isolation to insertion loss ratios of 60 dB : 1 dB are possible with such an arrangement. Because the r.f. energy is dissipated as heat in the resistance card, this type of attenuator is essentially limited to a few watts.

Bowness[1] and Reggia[2] have previously noted that the loss phenomenon is due to the transfer of r.f. energy between the incident polarization and another field polarized perpendicularly to it because of the transfer effect of the off-diagonal component of the tensor permeability of the ferrite material; the perpendicularly polarized field encounters a large attenuation constant due to the resistance card. Miller's[3] equations for coupled waveguides have previously been re-written to describe an elliptical Faraday rotator.[4] In this paper we extend this work for the case when the coupled waves have unequal attenuation constants.

It will be observed that the amplitude of the incident polarization is reciprocal, and has the same form as that given by Miller. This is in keeping with the experimental data. Hence, Miller's results with regards to the incident polarization remain unchanged. This gives a useful insight into the effect of unequal propagation constants on the attenuation. The perpendicular polarization, on the other hand, is non-reciprocal.

When the attenuation constant of the perpendicular polarization is greater than the incident one, and the phase velocities are the same, the periodicity in the power transfer between the coupled waves disappears for the ratio of attenuation difference/transfer effect more negative than -2, while for the same ratio more negative than -100 the power transfer has largely disappeared and the r.f. energy travels in the incident polarization with negligible attenuation. The attenuation difference/transfer effect ratio can be varied over large values by the d.c. magnetic field. When the phase velocities as well as the attenuation constants between the coupled waves are different, the periodicity between the coupled waves still disappears but now the maximum attenuation of the incident polarization is much reduced.

The paper concludes with some experimental data.

2. Coupled Waves in Magnetized Ferrite Section

For a system of vertically and horizontally polarized waves in a rectangular waveguide section coupled through the off-diagonal component of the tensor permeability due to a magnetized ferrite as shown in Fig. 1, we have

$$\frac{dE_x^{\pm}}{dz} = -\Gamma_x E_x^{\pm} \mp K_{yx} E_y^{\pm} \qquad \text{.....(1)}$$

$$\frac{dE_y^{\pm}}{dz} = \pm K_{xy} E_x^{\pm} - \Gamma_y E_y^{\pm} \qquad \text{.....(2)}$$

in which Γ_x, Γ_y are the perturbed complex propagation constants in the $x-y$ directions, i.e.

$$\Gamma_x = \alpha_x + j\beta_x, \qquad \Gamma_y = \alpha_y + j\beta_y$$

† Microwave Associates, Inc., Burlington, Massachusetts, U.S.A.

E_x^+, E_y^+ are the complex electric wave amplitudes in the $x-y$ directions when the applied d.c. magnetic field is in the same direction as that of r.f. propagation.

E_x^-, E_y^- are the complex electric wave amplitudes in the $x-y$ directions when the applied d.c. magnetic field is in the direction opposite to that of r.f. propagation.

$+K_{xy}$, $-K_{yx}$ are real quantities and represent the transfer effects of the off-diagonal component of the tensor permeability for the applied d.c. magnetic field in the same direction as that of r.f. propagation.

$-K_{xy}$, $+K_{yx}$ are real quantities and represent the transfer effects for the applied d.c. magnetic field in the direction opposite to that of r.f. propagation.

The form of the transfer effect has been commented on by Fox[5] but is not mandatory for what follows.

The solution to eqns. (1) and (2) are of the form

$$E_x^{\pm}|_{Kz} = A^{\pm}\exp(r_1^{\pm}z)+B^{\pm}\exp(r_2^{\pm}z) \qquad \text{.....(3)}$$

$$E_y^{\pm}|_{Kz} = C^{\pm}\exp(r_1^{\pm}z)+D^{\pm}\exp(r_2^{\pm}z) \qquad \text{.....(4)}$$

where

$$r_1^{\pm} = -\left(\frac{\Gamma_x+\Gamma_y}{2}\right) \pm jK\sqrt{\frac{(\Gamma_x-\Gamma_y)^2}{-4K^2}+1} \qquad (5)$$

$$r_2^{\pm} = -\left(\frac{\Gamma_x+\Gamma_y}{2}\right) \mp jK\sqrt{\frac{(\Gamma_x-\Gamma_y)^2}{-4K^2}+1} \qquad (6)$$

where we have written

$$|K_{xy}| = |K_{yx}| = |K|$$

To obtain the arbitrary constants we let $E_x^{\pm}|_0 = 1$ and $E_y^{\pm}|_0 = 0$ in eqns. (3) and (4). We also note that $\left.\frac{dE_x^{\pm}}{dz}\right|_{0+} = -\Gamma_x$, $\left.\frac{dE_y^{\pm}}{dz}\right|_{0+} = \pm K$ and $\Gamma_x > \Gamma_y$

When we do this we have

$$E_x^{\pm}|_{Kz} = \cos\left[\sqrt{\frac{(\Gamma_x-\Gamma_y)^2}{-4K^2}+1}.Kz\right].\exp\left[-\tfrac{1}{2}(\Gamma_x+\Gamma_y)z\right] - \frac{\frac{(\Gamma_x-\Gamma_y)}{2K}}{\sqrt{\frac{(\Gamma_x-\Gamma_y)^2}{-4K^2}+1}}.\sin\left[\sqrt{\frac{(\Gamma_x-\Gamma_y)}{-4K^2}+1}.Kz\right].\exp\left[-\tfrac{1}{2}(\Gamma_x+\Gamma_y)z\right] \qquad \text{.....(7)}$$

$$E_y^{\pm}|_{Kz} = \pm\frac{\sin\left[\sqrt{\frac{(\Gamma_x-\Gamma_y)^2}{-4K^2}+1}.Kz\right]}{\sqrt{\frac{(\Gamma_x-\Gamma_y)^2}{-4K^2}+1}}.\exp\left[-\tfrac{1}{2}(\Gamma_x+\Gamma_y)z\right] \qquad \text{.....(8)}$$

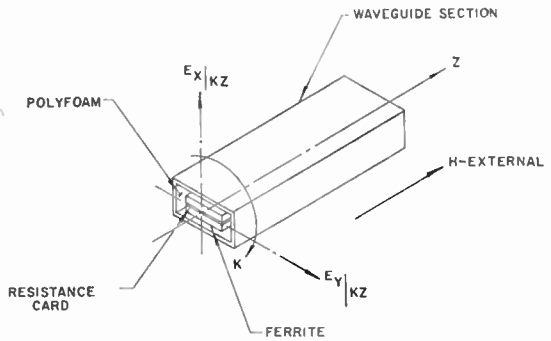


Fig. 1. Rectangular waveguide ferrite modulator.

From these equations we observe that the incident polarization is reciprocal notwithstanding the non-reciprocal character of the transfer effect. This is in keeping with the experimental data obtained on the absorption attenuator. The perpendicular polarization, on the other hand, is not reciprocal. Hence Miller's results for the incident polarization remain unchanged. This allows us to obtain significant insight into the effect of the ratio of unequal propagation constants to transfer effect.

3. Effect of Unequal Attenuation Constants on the Attenuation of the Incident Polarization for Equal Phase Velocities

To obtain this we write $(\Gamma_x-\Gamma_y)/K = (\alpha_x-\alpha_y)/K$ and $\beta_x = \beta_y$ in eqn. (7), and consistant with the physical structure of the absorption attenuator we evaluate eqn. (7) for negative values of $(\alpha_x-\alpha_y)/K$, i.e. the attenuation of the incident polarization is lower than that of the perpendicular one. In Fig. 2, the variation of $|E_x|$ is plotted against Kz for parametric values of $(\alpha_x-\alpha_y)/K$. This graph is reproduced from Miller's paper with slight changes in notation. From the latter graph we see that the periodicity is completely removed for $(\alpha_x-\alpha_y)/K = -2$, but the loss to $|E_x|$

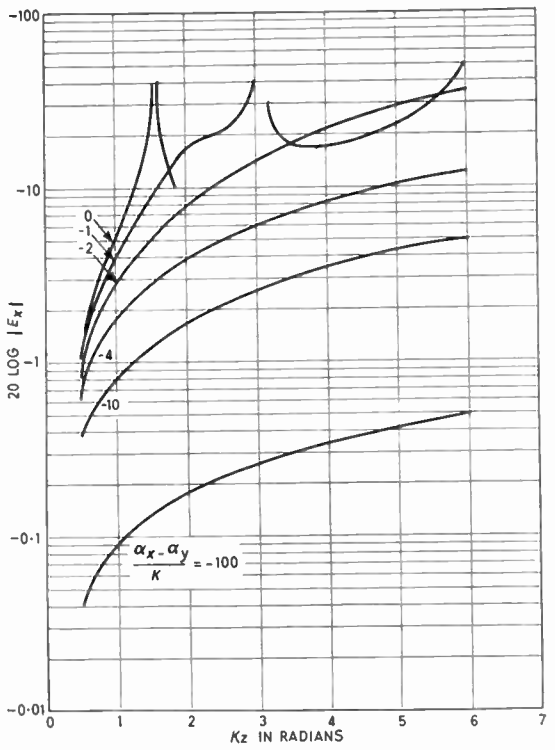


Fig. 2. Attenuation of incident polarization versus Kz for $(\beta_z-\beta_y)/K = 0$ and parametric values of $(\alpha_z-\alpha_y)/K$. The curve for $(\alpha_z-\alpha_y)/K = 0$ is periodic.

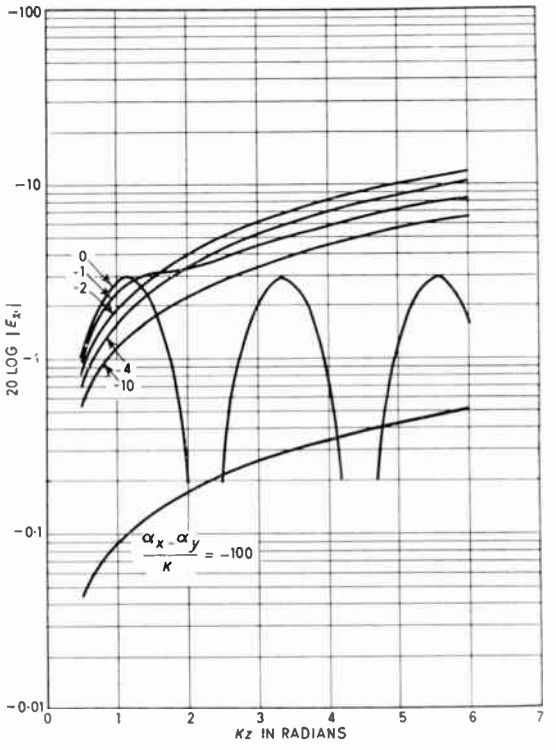


Fig. 3. Attenuation of incident polarization versus Kz for $(\beta_z-\beta_y)/K = +2$ and parametric values of $(\alpha_z-\alpha_y)/K$. The curve for $(\alpha_z-\alpha_y)/K = 0$ is periodic.

remains. On the other hand, for $(\alpha_x-\alpha_y)/K = -100$ the loss to the incident polarization is negligible. For a non-magnetized ferrite the transfer effect is zero and hence the last ratio is immediately obtained. The former ratio is obtained by adjusting the surface resistivity (ohms/square) of the resistance card in the magnetized state.

We also notice that the loss to the incident polarization is an increasing function of Kz, hence large dynamic values of attenuation are possible for arbitrary lengths of the ferrite slabs.

4. Effect of Unequal Attenuation Constants on the Attenuation of the Incident Polarization for Unequal Phase Velocities

In Fig. 3, we have replotted the attenuation of the incident polarization for $(\beta_x-\beta_y)/K = +2$ and parametric values of $(\alpha_x-\alpha_y)/K$. This is of practical interest, since for a rectangular waveguide loaded with ferrite material the former quantity is not zero. For this situation the maximum loss to the incident polarization is much reduced. This is in keeping with the experimental data which indicate that the maximum loss occurs when the ferrite shape behaves as a dielectric waveguide. However, the maximum attenuation curve still coincides with $(\alpha_x-\alpha_y)/K = -2$.

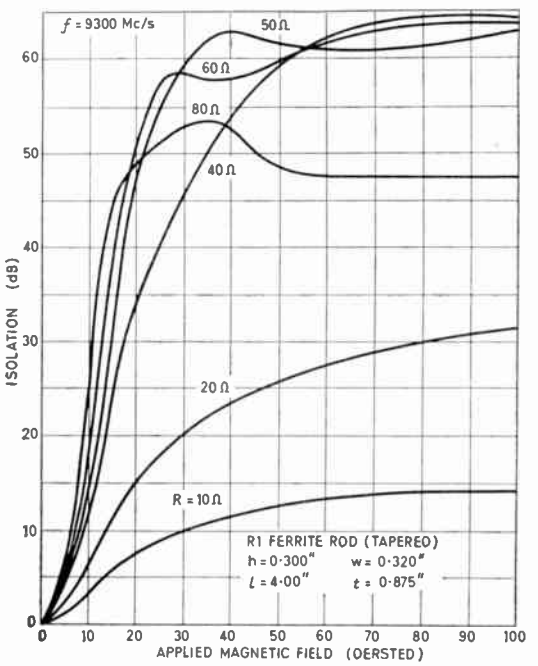


Fig. 4. Experimental attenuation of incident polarization versus applied magnetic field for various resistance cards.

5. Experimental Results

In Fig. 4 the attenuation of such a ferrite modulator versus applied magnetic field for different values of resistance cards is reproduced from Reggia's paper[2]. These curves were obtained at 9300 Mc/s for a ferrite shape 4 in long with $\frac{7}{8}$ in long tapers to give a good impedance match into the empty rectangular waveguide, the cross-section of the ferrite being 0·300 in × 0·320 in. The ferrite used was Ferramic R-1, a commercially available magnesium manganese ferrite.

Figure 4 shows the effect of the quantity $(\alpha_x-\alpha_y)/K$ on the attenuation of the incident polarization. The kinks in the attenuation curves for the cards above 40 ohms/square correspond to the quantity $(\alpha_x-\alpha_y)/K$ larger than -2, while the curve for the 40 ohms/square card correspond to $(\alpha_x-\alpha_y)/K = -2$.

6. Circulation using Coupled Waveguides with Unequal Attenuation Constants

A waveguide circulator using coupled waveguides in which one of the waveguides is loaded with a non-reciprocal ferrite phase shifter has previously been described.[5] If instead of non-reciprocal phase shift, we introduce non-reciprocal loss in the secondary waveguide with the help of a resonance isolator, we also have a circulator. In the forward direction, we let the isolator in the secondary waveguide be in its isolation condition, so that $(\alpha_x-\alpha_y)/K$ is large and the r.f. energy emerges in the primary waveguide. In the reverse direction, the isolator is in its insertion loss condition, and the r.f. energy is coupled into the secondary waveguide.

7. Conclusions

The absorption-type attenuator has been described in terms of coupled-wave theory. This approach has given useful insight into the effect of $(\alpha_x-\alpha_y)/K$ and $(\beta_x-\beta_y)/K$ on the attenuation of the incident polarization. The use of coupled waves to describe the absorption-type ferrite attenuator is also useful in that it does not restrict the latter to the geometry described, but provides for a broad family of possible geometries.

8. Acknowledgment

The author wishes to thank Microwave Associates, Inc., for permission to publish this paper.

9. References

1. C. Bowness, J. Q. Owen and K. M. Thomassen, "A low power amplitude modulator at microwave frequencies", Paper presented at I.R.E. Convention, Toronto, Canada, October 1957.
2. F. Reggia, "A new broad-band absorption modulator for rapid switching of microwave power", *I.R.E. Trans. on Microwave Theory and Techniques*, MTT-9, No. 4, pp. 343–9, July 1961.
3. S. E. Miller, "Coupled wave theory and waveguide applications", *Bell Syst. Tech. J.*, **33**, No. 1, pp. 661–719, May 1954.
4. J. Helszajn, "A novel ferrite quarter-plate", *The Radio and Electronic Engineer*, **27**, No. 6, pp. 455–8, June 1964.
5. A. G. Fox, S. E. Miller and M. T. Weiss, "Behavior and application of ferrites in the microwave region", *Bell Syst. Tech. J.*, **34**, No. 1, pp. 5–103, January 1955.

Manuscript first received by the Institution on 14th April 1964 and in final form on 21st September 1964. (Paper No. 963.)