

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the British Institution of 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 26

OCTOBER 1963

NUMBER 4

CHANGE OF THE INSTITUTION'S NAME

Special General Meeting to precede Annual General Meeting

At a meeting of the Council of the Institution held on Thursday, 10th October, 1963, it was unanimously resolved that a Special General Meeting of Corporate Members be called to confirm the Resolution of the Council to alter and amend the Institution's Royal Charter of Incorporation which, when allowed by Her Majesty in Council, will give effect to a change in the title of the Institution. Thus the proposal already tacitly accepted by members all over the world that the Institution's title should more accurately indicate the professional activities of its members will be implemented.

Just a year ago an announcement was made in the Institution's *Journal* of the Council's intention to change the name of this publication to *The Radio and Electronic Engineer*. Comment was invited from members on the even more significant matter of making an alteration to the name of the Institution itself—often mooted during recent years. The various proposals were reviewed and at a dinner of the Council and Committees held in London on 27th November, 1962 (reported in *Proc. Brit. I.R.E.*, January/February 1963), the President announced that the approval of the Privy Council was being sought to change the Institution's name to "The Institution of Electronic and Radio Engineers". Lord Mountbatten continued: "The word 'electronic' has become one of the most widely used words in our language and much of the activity of our Institution covers purely electronic subjects as opposed to radio. Our activities are spreading far beyond the border of Great Britain, and I think it wrong that we should restrict our title by embodying the word 'British' which is not employed in the title of any similar chartered engineering institution".

As readers of a "A Twentieth Century Professional Institution"* will know, the inclusion of the word 'electronic' in the title of the Institution is not a new idea. It was, in fact, ventilated in the Institution in 1927! It is, however, only in the last two decades that industrial development has exploited the possibilities revealed in the course of advancing radio science.

It is now appropriate, therefore, that more adequate description be given to the professional activities of members of the Institution. The Special General Meeting, to be held just prior to the Annual General Meeting of the Institution on 27th November, represents a very natural step in the future development of Institution activities.

Formal notice giving precise details of the Resolutions required is being despatched separately to all Corporate Members.

* "A Twentieth Century Professional Institution—The Story of the Brit. I.R.E.", published by the Institution, price 30s.

INSTITUTION NOTICES

I.E.E. President's Inaugural Address

In his Inaugural Address as President of the Institution of Electrical Engineers, Sir Albert Mumford, K.B.E., Engineer-in-Chief of the General Post Office, referred to the benefits which were to be derived through co-operation between engineers. He especially stressed that the inauguration of the Engineering Institutions Joint Council in October 1962 offered opportunities of a fuller and more effective collaboration between Institutions.†

Sir Albert continued: "In some fields, collaboration has developed and will continue in a more personal way. I have in mind, of course, collaboration between the British Institution of Radio Engineers and ourselves, expressed, for instance, in our joint activities in the educational, computer and medical-electronics fields and in the co-sponsorship with them and with the interested American societies, of the highly successful International Telemetry Conference which has just been held in London. I am pleased to say that this collaboration with our sister Institution, whose field of activity we share is to be seen not only in London but in our local Centres also".

Introducing his main subject of "Communications in the Public Service of the United Kingdom", Sir Albert outlined the increasing trend towards full automation of the public telephone service, and described developments in the transmission field that would provide more circuits at lower costs. These included microwave radio links in the trunk network and, eventually, long distance waveguide transmissions. He discussed the demands of television and how these demands were integrated into the other public services in Great Britain, and then dealt with the problems of worldwide telecommunications, including the special role of satellites.

Radio and Electronics Research in Great Britain

The Survey prepared at the request of the President by the Institution's Research Committee on "Radio and Electronics Research in Great Britain" has now been published in the July/August issue of the *Proceedings of the Brit.I.R.E.* (Vol. 1, No. 5). Members in the United Kingdom will have automatically received their copies of this issue. Members overseas may obtain copies of the issue price 3s. each; the cost to non-members is 7s. 6d. per copy.

The Survey deals with the role played in radio and electronics research by various government departments, universities and the radio and electronics

† Reference to the formation of the Engineering Institution's Joint Council was made in the November 1962 issue of the *Brit. I.R.E. Journal*. The founder members of the Council comprise the 13 major engineering institutions in Great Britain, including the Institution of Electrical Engineers and the British Institution of Radio Engineers.

industry itself. Its conclusions urge the establishment of a research council which would comprise representatives of these three spheres of activity and it is encouraging to record that preliminary steps are now being taken to set up this council with full legal status.

Appendices to the Survey analyse the radio and electronics research programmes of universities and colleges of technology in Great Britain, recording the subjects of the projects and the number of research workers employed. Details are also given in rather more general terms of research in Government Departments, the United Kingdom Atomic Energy Authority, the British Broadcasting Corporation and the General Post Office, and in Industry.

Fondation Européenne de la Culture

A conference of engineering students and young technologists from Universities and Colleges of Technology throughout Western Europe was held at the University of Grenoble in September. The Engineering Institutions Joint Council was invited to send representatives and the Institution nominated Mr. R. D. Pringle (Student), who was awarded the first Mountbatten Research Studentship in 1962.

Lectures by well-known scientists and engineers took place each morning and in the afternoons the delegates were divided into four groups which discussed:

- (a) participation of the engineer in administration;
- (b) team-work and external relations with other companies and countries;
- (c) the education of the engineer;
- (d) the engineer and conflicts between social classes, different countries and regions.

Mr. Pringle was particularly concerned with the group discussing education.

Mr. Pringle reports that, in his view, the Conference achieved more by informal meetings of delegates from all over Europe and the discussion of their problems rather than through the resolutions passed by the discussion groups. A summary of these resolutions will be available by about the beginning of November, and Mr. Pringle has kindly offered to send copies to any members of the Institution who are interested.

Correction

The caption to Fig. 1 of the article "The N.I.R.N.S. 7 GeV proton synchrotron" on page 203 of the September issue of *The Radio and Electronic Engineer* should be amended to read "... showing installation of the beam extraction box".

Automatic Control of Billet Mill Saw Measuring Stops

By
G. COOPER†

Presented at the Convention on "Electronics and Productivity" in Southampton on 19th April 1963.

Summary: Measurement of the measuring stop position is by an electromagnetic digitizer and the control system is built-up from standard plug-in transistor logic units. The drive of the measuring stop is by a two speed pole amplitude modulated type pole-changing induction motor. Use is made of a fast traverse, slow approach cycle, and the final approach is always from one direction to eliminate back-lash errors. A pneumatic brake is used to bring the stop to rest at the final position. A repeatable accuracy of better than $\frac{1}{16}$ in has been obtained on a stop capable of measuring length of up to 60 ft.

1. Introduction

Increasing attention is being currently focused on the automation of the cutting up of rolled steel products immediately after leaving the mill. The benefits sought are the reduction of cost of the cutting operation and particularly the optimization of cutting to reduce the incidence of unusable short remainders.

In general, progress towards the fully automated cutting plant is being made step-by-step, and this paper describes a position control system using digital electronics for the saw measuring stops. The system is

this type of stop is particularly suitable for high production automated plants. The positioning control was devised with the object of providing a rugged, reliable system capable of standing along or of fitting into a fully automated plant.‡ The digital techniques employed have been described in connection with other steel industry position controls,§ the same range of standard logic units being employed. The emphasis of the design of this control has been on economy, ruggedness, and flexibility together with a performance which will be maintained for many years of service.

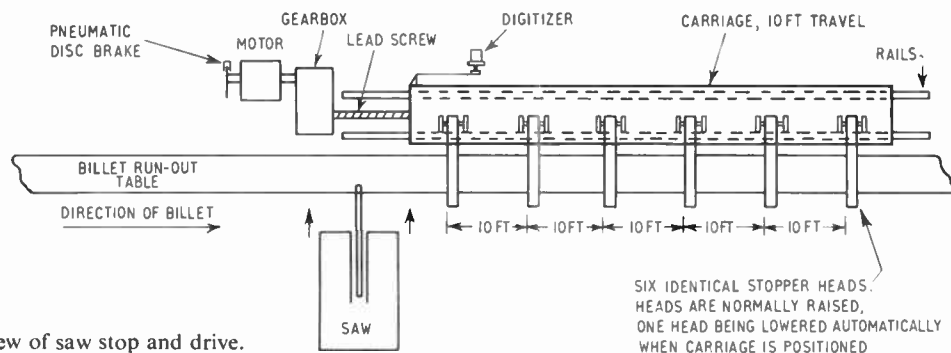


Fig. 1.
Plan view of saw stop and drive.

therefore immediately suitable for integration into a completely automatic computer controlled plant.

Figure 1 shows diagrammatically the construction of a multiple head saw stop. Typically the carriage has a travel of 10 ft, with retractable heads at 10 ft centres. By the use of 6 heads, settings of up to 60 ft can be made with a much shorter average positioning time than with a single head and a sixty foot travel. Hence

2. Drive System

The drive of the carriage is by means of a lead screw and nut, the screw being driven through gearing from an electric motor. A pneumatic brake is mounted on a rear shaft extension of the motor.

The obvious drive system for this type of high inertia positioning duty would be a d.c. motor and

‡ J. Clyne, "Automatic Control of Cutting up of Steel Billets" B.I.S.R.A. report No. PE/R57/62, October 1962.

§ W. Gregson and G. Cooper, "Automatic control of hydraulic forging presses", *Control*, 5, page 99, June 1962.

† Lancashire Dynamo Electronic Products Ltd., Rugeley, Staffordshire.

Ward-Leonard set. Many steel companies, however, prefer to use a.c. squirrel cage motors where possible, because of the reliability and freedom from maintenance of this type of machine. For this reason, and the important one of greatly reduced cost, a 35/8.75 hp, 1350/335 rev/min, 4/16 pole two-speed a.c. machine was adopted for the prototype system.

The machine used is of some interest, since it employs the relatively recently developed pole amplitude modulated winding. This winding uses the same conductors in different connections for both high and low speed operation. For this start-stop application, one of the major attractions is that of reduced inertia within standard frame dimensions, since the windings can be fitted into a smaller frame size than the conventional double-wound machine. Hence a smaller diameter rotor can be used. Important savings in cost and weight also result. For the particular machine used, the savings are tabulated below:

	Conventional Machine	Pole Amplitude Modulated Machine
Inertia	100%	78%
Cost	100%	77%
Weight	100%	83%

The high speed winding is used to traverse the stop at 30 ft/min to within 6 in of required final position, when the low speed winding is energized giving regenerative braking down to a stable speed of 7.5 ft/min. If the stop is approaching the saw, the slow speed run-in is followed by application of the brake at the required position. If the stop is moving away from the

saw, the carriage is allowed to overshoot and is then plug-reversed on the low speed winding, and runs into its final position towards the saw. This system ensures that all mechanical back-lash is taken up against the direction of impact of an incoming billet. Also, overrun as the brake is applied is unimportant so long as it is consistent. The positioning cycle is described in more detail in a later section.

3. Control Elements

The positioning system is of the true-binary coded digital type. That is, all positions are expressed numerically in binary code, in this case as a 12 digit binary number, with a least significant digit of $\frac{1}{32}$ in.

The example below shows the method of number representation.

Digit	12	11	10	9	8	7	6	5	4	3	2	1
Value of Digit (in)	64	32	16	8	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$
Typical No.	0	1	1	0	0	0	1	0	0	1	0	0
Decimal Equivalent	32+16							+1			+ $\frac{1}{8}$ = 49 $\frac{1}{8}$ in.	

The numerical information is handled on transistor logic circuits, '0' being represented by zero signal. '1' by -6 V to -14 V. The advantages of handling information digitally are well known: freedom from drift, ease of storage and display, and in this application, compatibility with an instructing computer.

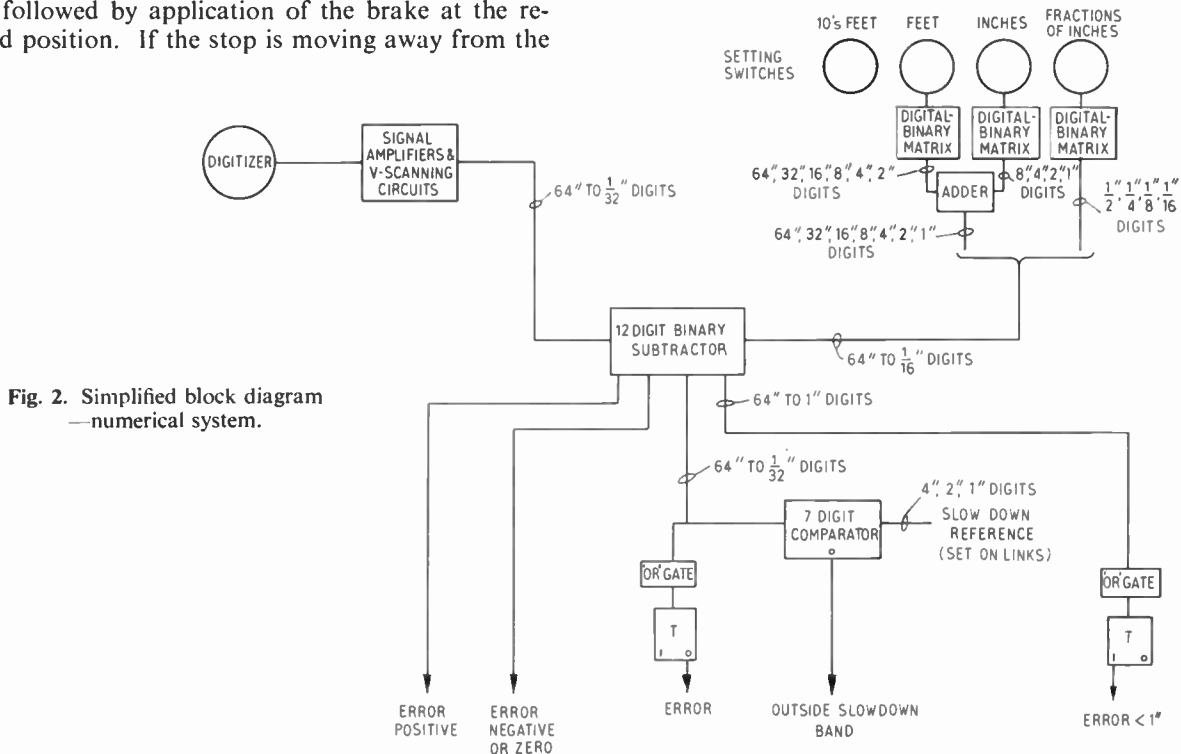


Fig. 2. Simplified block diagram —numerical system.

3.1. Carriage Position Measurement

3.1.1. Digitizer drive

The block diagram of the measurement system is shown on Fig. 2. The carriage position is measured by an electro-magnetic digitizer, driven mechanically via a wire cord uncoiled from a drum by the movement of the stop and recoiled by a return spring built into the digitizer. The output of the digitizer is 12 binary digits, giving a resolution of 1 part in 4096. The wire rope has been proved over some years of service in other applications to be robust, and satisfactorily accurate. It is much less costly than a rack and pinion, and is free from back-lash.

3.1.2. Principle of digitizer operation

The digitizer was originally developed for forging press position controls. This application, in common with the saw stop application, involved high temperature, vibration, shock, and heavy wear and tear. For this reason both the contact and photoelectric types of digitizer were considered insufficiently trouble free, and a digitizer free from moving contacts and light sources was developed.

The binary pattern is defined by arc-shaped magnets cast into an epoxy-resin disc (see Fig. 3). The discs are rotated by the motion of the carriage. Each magnetic track passes over two small ferrite pick-off cores mounted just below the surface of a fixed epoxy disc. A 0.020-in air-gap is maintained between fixed and moving discs. Each core has a multi-turn output winding, and is threaded by a single-turn primary winding excited at 70 kc/s. Thus, when no magnet is over the core, a 70 kc/s voltage is produced in the output winding. When a magnet is over the core, however, the magnetic flux keeps the core saturated and prevents coupling between the windings. In this case there is no signal in the output winding. The output-winding signal is rectified, and after amplification becomes a digital signal suitable for driving the numerical units: a magnet generates a '0' output, no magnet a '1' output.

In order to reduce the inertia of the digitizer, three discs are used, each generating four digits and geared together in the ratio 256 : 16 : 1. To avoid ambiguities when several tracks change together, and to allow for gearing tolerances, the well-known V-scanning method is used; starting with the least significant, (i.e. $\frac{1}{32}$ in) track, the signals select which of the two pick-offs in the next track is to be read, avoiding reading pick-offs near the end of a digit segment.

To simplify installation problems, the 70 kc/s driver-oscillator is mounted in the digitizer, as are the rectifiers for the output signals. Thus, both supply and signals are d.c., and the digitizer may be positioned up to 200 ft from the cubicle, and connected by

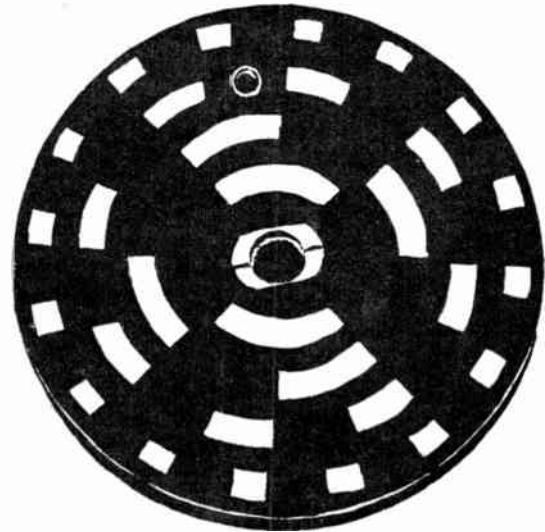


Fig. 3. Magnet disc from digitizer (2.5 in dia.).

conventional industrial wiring. Silicon semiconductors are used throughout, and the units are rated at 85° C ambient.

3.2. Required Position Setting

The length required to be cut is set manually on rotary switches. The switches used are of a stud type with a heavy wiping action to ensure trouble free operation. They are calibrated in tens of feet, feet, inches, and fractions of an inch. The tens of feet switch is used directly to select the stopper head required to be lowered, and the settings of the remaining three switches are converted to binary code and summed to give the reference position for the carriage position control. Conversion is by diode matrices, and the numerical additions are carried out in ADDER units built up from transistor/diode/resistor AND, OR and amplifier stages. The logic units are described in Section 3.4.

3.3. Error Measurement Circuits

The system error, i.e. the distance between the present position of the carriage and the required position, is calculated by a reversible subtractor, built up from logic units, which generates the sign and modulus of the error in binary form.

By comparing the modulus of the error with a pre-set 'slow down' distance in a numerical comparator, a further signal is produced indicating whether a carriage is within or outside the 'slow approach' zone. The 64 in to 1 in digits of the error are fed also to an OR gate, an output from which indicates an error greater than 1 in. Sufficient information in the form of on/off signals is now available to control the position sequence described in Section 5.

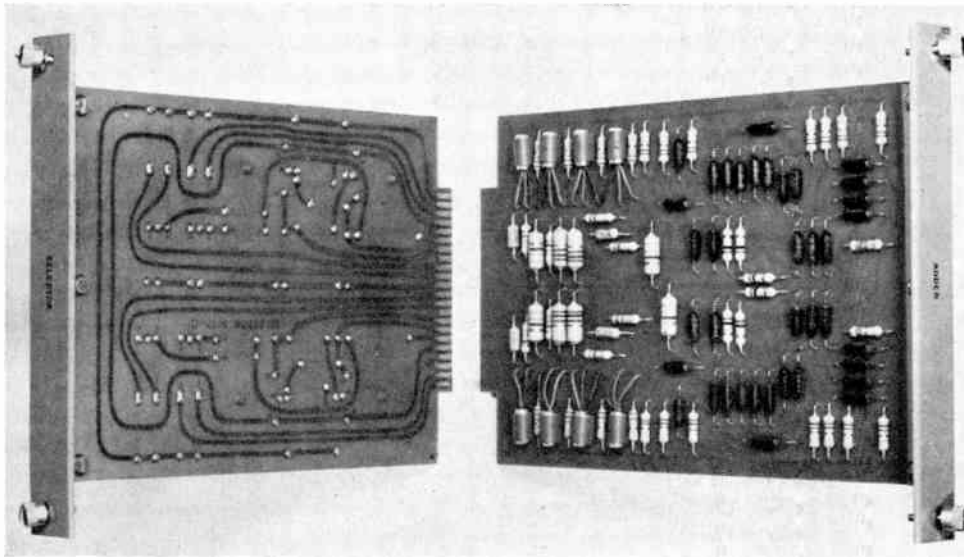


Fig. 4. Typical logic units.

3.4. Logic Unit Design

The logic units use supplies of ± 14 V d.c., un-stabilized, and derived without smoothing from a 6 phase double star connected transformer/rectifier unit.

The design tolerances are such that, for example, all transistors can fall to 70% of manufacturers

minimum gain (β'), with all resistors on limit tolerance in the worst direction, and with supplies with ± 2 V unbalance, before the unit will fail to operate. Germanium semiconductors are used throughout, the design temperature rating being 55° C.

All logic circuits are mounted on $8\frac{1}{2} \times 6$ in printed circuit panels, plugged into 22 way single side or 44 way double sided, 2×10^{-4} in hard gold plated, edge connector sockets. The printed circuit plug is similarly gold plated, and no service failures have been experienced due to contact troubles.

To utilize the printed circuit area efficiently, multiples of basic circuits are mounted on one unit, e.g. 12 separate OR gates, 8 separate AND gates, 5 separate logic amplifiers (triggers). Figure 4 shows a typical logic unit.

The units slide into pressed steel frames, mounting 12 units in a 'box'. The frames are designed to allow forced ventilation where necessary in high ambient temperatures.

Rapidity of rectification of a control system is almost as important as inherent reliability. A monitoring facility, consisting of a panel of 12 lamps with built-in driving amplifiers (in a 12 digit system), which can be plugged into all the key points of the system, e.g. both subtractor inputs and the subtractor output, is therefore provided.

With the aid of a simple block diagram, showing which units perform which functions, and the monitoring facility, a fault can rapidly be pinned down to, say, one of three units which can then be changed sequentially for serviceable spares in a few minutes.

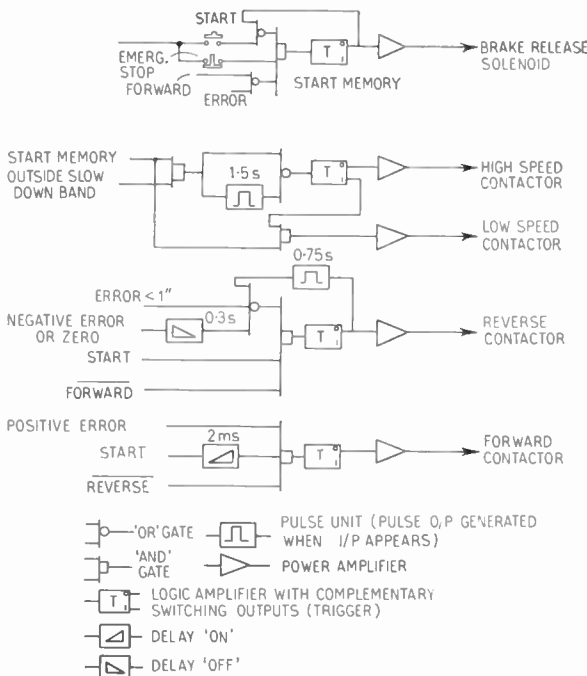


Fig. 5. Sequencing logic.

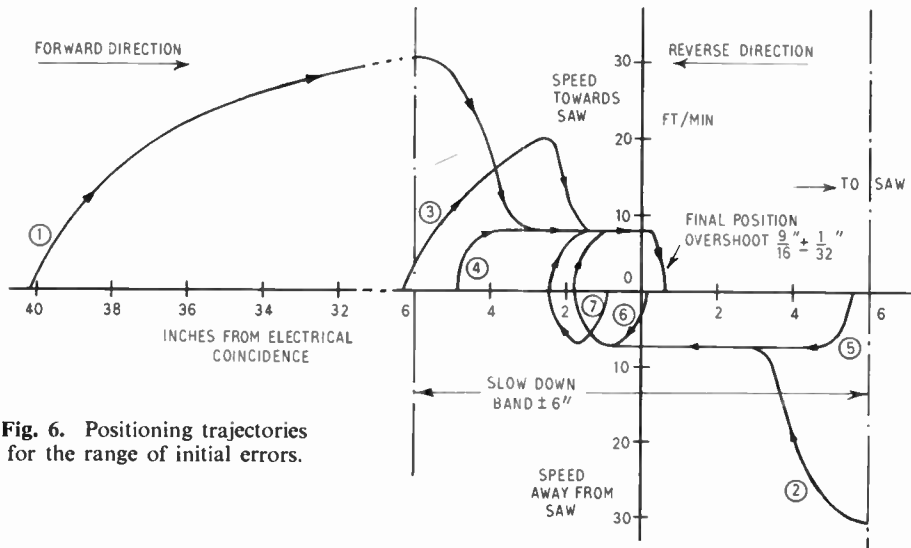


Fig. 6. Positioning trajectories for the range of initial errors.

4. Positioning Sequence

The accuracy of the system is entirely dependent upon the consistency of stopping distance after the de-energization of the motor and the simultaneous energization of the brake. The zero of the digitizer scale is displaced by this overrun from the actual mechanical zero of the system in order to bring the final position to the required point. It is therefore essential that when the 'stop' signal is given, the drive is running steadily towards the saw at the low speed of the motor. The positioning cycle is designed to ensure this whatever the initial error distance may be when the cycle starts. The control logic is shown in Fig. 5 and Fig. 6 shows the trajectories followed for the range of initial errors. The curves are constructed from measurements made on the prototype system.

4.1. Control Sequences

4.1.1. Large initial positive error

- (a) Motor starts in forward high speed, and runs carriage to 6 in from coincidence.
- (b) Motor is switched to forward low speed.
- (c) At coincidence, motor is de-energized and brake applied.

4.1.2. Large initial negative error

- (a) Motor starts in reverse high speed, and runs carriage to 6 in from coincidence.
- (b) Motor is switched to reverse low speed.
- (c) 0.3 seconds after reaching coincidence, motor is switched to forward slow speed.
- (d) At coincidence, motor is de-energized and brake applied.

4.1.3. Initial error just greater than 6 in positive

- (a) Motor starts in forward high speed, and is held in high speed for 1.5 seconds.

- (b) Motor is then switched to forward low speed.
- (c) At coincidence, motor is de-energized and brake applied.

4.1.4. Initial error between 6 in position and 1 in positive

- (a) Motor starts in forward low speed.
- (b) At coincidence, motor is de-energized and brake applied.

4.1.5. Initial error between zero and 6 in negative

- (a) Motor starts in reverse low speed.
- (b) 0.3 second after reaching coincidence, motor is switched to forward low speed.
- (c) At coincidence, motor is de-energized and brake applied.

4.1.6. Initial error between 1 in positive and zero

- (a) Motor starts in reverse low speed, and is held in reverse low speed for 0.75 seconds.
- (b) Motor is switched to forward low speed.
- (c) At coincidence, motor is de-energized and brake applied.

4.2. Accuracy

Random inaccuracy occurs due to variations in response time of the brake and in release time of the

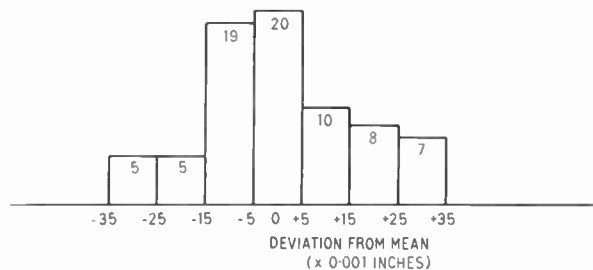


Fig. 7. Random positioning errors.

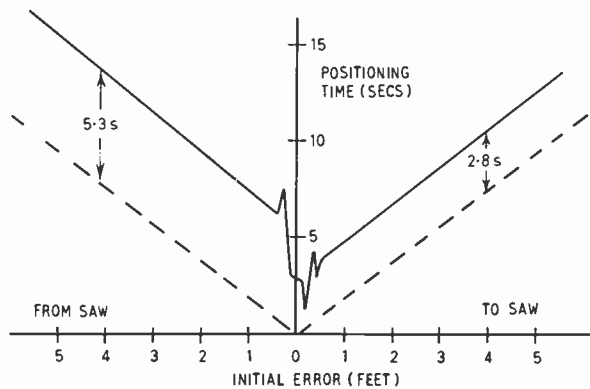


Fig. 8. Positioning time.

forward low speed contactor. Random error will also occur due to variations in the motor speed at the time of de-energization, resulting from the different approach trajectories followed from different initial errors. Figure 7 shows a series of 74 stops at the same nominal position. The initial errors were varied over the complete positive and negative range. The maximum measured deviation was 0.035 in, and 66% were within 0.015 in.

Longer term inaccuracies due to brake wear and mechanical wear can be simply taken up by adjusting the digitizer cable anchorage, thus moving the digitizer zero.

4.3. Positioning Speed

Figure 8 shows the positioning times for varying distances of initial error. The positioning time is

mainly a function of the maximum drive speed, and can be varied to a certain extent to suit user requirements by varying the motor rating. Where necessary the 'settling' time when moving away from the saw could be reduced by use of an extra comparator to give different slow-down points for cycles toward and away from the saw.

5. Conclusions

The system described is a satisfactory means of providing heavy duty position controls using a squirrel-cage induction motor drive.

The reliability experienced from the techniques used, in other applications, indicates a probable fault rate of not more than one fault per two years of service.

The control techniques are applicable to many other position control problems.

6. Acknowledgments

The project described was a co-operative venture, and the author would like to acknowledge in particular the contributions made by Mr. W. L. Goldie of Lamberton & Co. the saw manufacturers, Mr. W. D. Lewis of Lancashire Dynamo Electronic Products, Mr. H. Sterling of Lancashire Dynamo & Crypto, who designed and supplied the motor, and Mr. K. Yeomans of British Iron & Steel Research Association, who carried out much of the system analysis and performance measurement.

Manuscript received by the Institution on 19th March 1963. (Paper No. 850.)

© The British Institution of Radio Engineers, 1963

DISCUSSION

Under the Chairmanship of Mr. M. James

Mr. A. E. Z. Cohen: Referring to the digitizer, I observe that ambiguity is overcome by using 2 pick-offs per track. If a cyclic progressive code were used only one pick-off per track would be necessary. I assume that this has not been done because the subsequent logic becomes more complicated. Is this so?

The Author (in reply): In practice, the amount of logic required to convert from V-scan pick-off outputs to true binary does not differ greatly from that required to convert a cyclic progressive code to true binary.

The V-scan technique was used in the digitizer described since it permits the edge of a digit to be displaced by up to $\pm 25\%$ of the length of that digit from its true position without misreading occurring; on a cyclic code the error must be small as a percentage of length the least significant digit, and thus effectively a closer tolerance pattern is needed. It would not be easy to generate magnetic patterns to this accuracy.

Mr. J. A. Adams: The use of a wire digitizer drive I find interesting. Could the author indicate if any special traversing gear is necessary and also indicate the form of the take-up equipment?

The Author (in reply): No traversing gear is needed. Lateral adjustment of the cable anchorage is provided which permits a position to be found at which the cable coils neatly.

The take-up equipment is in the form of a pre-tensioned 'clock' type spring, which gives a satisfactorily flat tension/deflection curve.

Mr. R. L. Duthie: In Fig. 7 of the printed paper, random positioning errors are shown to extend over ± 35 thou. The lecturer says that the digitizer errors are $\pm \frac{1}{32}$ in. Does this leave only ± 3 thou. for braking inconsistencies, or are these added respectively. Does this not conflict with the spread of settling point of $\pm \frac{1}{32}$ in shown in Fig. 6?

The Author (in reply): The errors of $\pm \frac{1}{32}$ in given in Figs. 6 and 7 are around the 'electrical coincidence' of the system. To this error must be added the $\pm \frac{1}{32}$ in by which the digitizer mechanical drive error may cause the electrical coincidence to differ from the actual measurement required. Taken together, these errors add up to $\pm \frac{1}{16}$ in overall error limit claimed in the summary to the paper.

Optimum Processing for Acoustic Arrays

By

WOUTER VANDERKULK,

Ph.D. †

Presented at the Symposium on "Sonar Systems" in Birmingham on 9th–11th July 1962.

Summary: Bryn's optimum detector is briefly described. Its performance is compared with the conventional detector for background noise consisting of self-noise and isotropic sea-noise.

1. Introduction

In a recent paper, ‡ F. Bryn has developed a method for optimum detection of stationary small-amplitude sound waves received by a K -element array in the presence of stationary Gaussian noise. His detector is an approximation to the Neyman-Pearson likelihood ratio. As a consequence of the above assumptions concerning signal and noise, the optimum detector is quadratic and can also be characterized as that quadratic detector, W , which maximizes the conventional output signal-to-noise ratio (defined as the ratio of the mean value of W when signal alone is present to the standard deviation of W when noise alone is present).

The design of the optimum detector presupposes complete knowledge of the cross-spectra of the signal to be detected and of those of the noise processes at the array elements. In contrast, the conventional detector only requires the relative arrival times of the received signals in order to implement their in-phase addition.

Bryn's method is perhaps most useful for array and processing design when signal or noise possess unusual features (e.g. when the noise field has pronounced directional characteristics). The present note, however, is concerned solely with the comparison of optimum and conventional detection when so-called 'normal' conditions prevail, i.e. when the following requirements are satisfied:

- (1) The medium is homogeneous at the array.
- (2) The signal presents a single, perfectly coherent planar wavefront to the K -element array.
- (3) The sea-noise is isotropic and is generated by uniformly distributed planar wave-fronts impinging on the array.

- (4) The self-noise is homogeneous and is statistically independent of the sea-noise. There exists statistical independence between the self-noise processes at different elements.

These assumptions are in addition to those postulated by Bryn and formulated at the beginning of this section.

The criterion used in comparing optimum with conventional detection has been the optimum versus the conventional *array gain* (or *directivity factor*), taking the effect of self-noise into account. Some attention has also been paid to the beam-pattern, particularly to the width of the main lobe.

It is well-known that the conventional detector is practically optimum when the average separation between neighbouring elements is of the order of half of one wavelength or more. The present note deals with the situation where this separation is small compared to the wavelength. In that case, the optimum array gain can significantly exceed the conventional array gain. In fact, the average optimum array gain of a spherically symmetric array with a *fixed* radius-to-wavelength ratio tends to infinity as the number of elements of the array is increased indefinitely. This is true even when self-noise is present. Such behaviour is in sharp contrast to that of the conventional array gain which remains bounded when the radius-to-wavelength ratio is fixed.

Unfortunately, owing to the assumed statistical independence of the self-noise processes at different elements, the growth of the optimum gain with the number of array elements is rather slow when self-noise is present. For example, when the radius-to-wavelength ratio is two, and the self-noise is 20 dB below the sea-noise, about 400 elements suffice to achieve a conventional array-gain of 21 dB. However, some 1 000 000 000 elements are required to obtain an optimum gain which is only 6 dB higher, i.e. which is 27 dB. Little practical benefit appears to be gained from the use of the optimum detector whenever the maximum array dimension is in the order of a few wavelengths or more, particularly in view of the rather complex electronics needed to implement it. The situation seems somewhat more favourable for the

† University of California, San Diego, Marine Physical Laboratory of the Scripps Institution of Oceanography, San Diego 52, California, and International Business Machines Corporation, Owego, New York. This work was performed while on a leave of absence at the Scripps Institution of Oceanography.

‡ Finn Bryn, "Optimum signal processing of three-dimensional arrays operating on Gaussian signals and noise", *J. Acoust. Soc. Amer.*, 34, No. 3, pp. 289–97, March 1962.

optimum detector when the number of array elements is small (around twenty, say) and the maximum array dimension is about one wavelength or less.

In the next few sections, these matters are discussed in more detail for the spherically symmetric array, the linear array, and the ring of equally spaced elements.

2. Formulae for Optimum Array Gain and Beam Pattern

In this section, mathematical expressions for the array gain and beam pattern of the optimum detector are presented. The reader is referred to Bryn's paper† for a detailed derivation.

In accordance with the 'normal' conditions postulated in the preceding section, we restrict our attention to planar wavefronts. Such a wavefront is uniquely defined by the unit vector, u , normal to the wavefront and in the direction of its travel. The relative phases with which this wavefront is received by the elements of a K -element array are uniquely described by the K complex phase-factors

$$p_k(u, \lambda) = \exp[2\pi(x_k \cdot u)j/\lambda]; \quad k = 1, \dots, K \quad \dots\dots(1)$$

where x_k is the coordinate vector of the k th element with respect to some orthogonal coordinate system.

Since the sea-noise is assumed to be generated by uniformly distributed planar wavefronts, it follows that the normalized sea-noise cross spectrum, $q_{k,h}$, between the k th and the h th array elements is given by

$$q_{k,h}(\lambda) = \langle p_k(u, \lambda)p_h(u, \lambda)^* \rangle = \frac{\sin(2\pi d/\lambda)}{2\pi d/\lambda} \quad \dots\dots(2)$$

where $\langle \rangle$ denotes the uniformly weighted average over all unit vectors u and where d is the distance between the k th and h th elements. The symbol z^* represents the complex conjugate of z .

$G_o(u, \lambda)$ shall denote the optimum array gain (or directivity factor) when the array is steered in the direction of the unit vector u . Similarly, $G_c(u, \lambda)$ denotes the conventional array gain. It can be shown that†

$$G_o(u, \lambda) = \sum_{k=1}^K Z_k p_k(u, \lambda) \quad \dots\dots(3)$$

where Z_k ; $k = 1, \dots, K$ is the solution of the K linear equations

$$\sum_{k=1}^K Z_k q_{k,h}(\lambda) + \xi(\lambda)Z_h = p_h(u, \lambda)^*; \quad h = 1, \dots, K \quad \dots\dots(4)$$

in which $\xi(\lambda)$ is the ratio of the self-noise power to the sea-noise power for the wavelength λ . According to the assumptions made in the preceding section, this ratio is the same for all K elements of the array.

† F. Bryn, *loc. cit.*

Likewise,

$$\frac{1}{G_c(u, \lambda)} = \frac{1}{K^2} \sum_{k,h=1}^K p_k(u, \lambda)^* q_{k,h}(\lambda) p_h(u, \lambda) + \frac{\xi(\lambda)}{K} \quad \dots\dots(5)$$

The optimum detector's (power) beam pattern shall be denoted by $\beta_o(v, u, \lambda)$ where, as before, u defines the direction in which the array is steered while v is the unit vector in the direction of the received signal. Bryn has shown that

$$\beta_o(v, u, \lambda) = \left| \frac{1}{G_o(u, \lambda)} \sum_{k=1}^K Z_k p_k(v, \lambda) \right|^2 \quad \dots\dots(6)$$

Likewise, the conventional power beam pattern is given by

$$\beta_c(v, u, \lambda) = \left| \frac{1}{K} \sum_{k=1}^K p_k(u, \lambda)^* p_k(v, \lambda) \right|^2 \quad \dots\dots(7)$$

The Z_k 's defined by eqn. (4) are the transfer functions of K filters which constitute the major portion of the optimum detector and through which the outputs of the K array elements must be passed prior to their addition. The computation of the Z_k 's from eqn. (4) represents the most complicated step in the calculation of G_o and β_o from eqns. (3) and (6). In this regard, it is noted that the $q_{k,h}$ -matrix is real and symmetric according to eqn. (2). This property is due to the rather special sea-noise model adopted in Section 1. In general, the $q_{k,h}$ -matrix is not real although it is always Hermitean (i.e. $q_{h,k} = q_{k,h}^*$).

An explicit and formally simple expression for Z_k exists in terms of the eigenvalues and eigenvectors of the $q_{k,h}$ matrix. Denote the K eigenvalues by Kv_1, \dots, Kv_K and let $(E_{1,m}, \dots, E_{K,m})$ represent the eigenvector corresponding to the eigenvalue Kv_m . The eigenvectors are assumed to be orthonormal in the Hermitean sense (i.e. $\sum_k E_{k,n} E_{k,m}^*$ is zero when $n \neq m$ and unity when $n = m$; $n, m = 1, \dots, K$). Then

$$Z_k = \sum_{m=1}^K D_m(u)^* E_{k,m} / (Kv_m + \xi(\lambda)) \quad \dots\dots(8)$$

where

$$D_m(u) = \sum_{k=1}^K p_k(u, \lambda) E_{k,m} \quad \dots\dots(9)$$

Hence,

$$G_o(u, \lambda) = \sum_{m=1}^K |D_m(u)|^2 / (Kv_m + \xi(\lambda)) \quad \dots\dots(10)$$

and

$$\frac{1}{G_c(u, \lambda)} = \frac{1}{K} \sum_{m=1}^K v_m |D_m(u)|^2 + \frac{\xi(\lambda)}{K} \quad \dots\dots(11)$$

Similarly,

$$\beta_o(v, u, \lambda) = \left| \frac{1}{G_o(u, \lambda)} \sum_{m=1}^K \frac{D_m(u)^* D_m(v)}{Kv_m + \xi(\lambda)} \right|^2 \quad \dots\dots(12)$$

We note that the sum of the v_m 's is unity and that the sum of the $|D_m(\mathbf{u})|^2$'s is K . The eigenvalues v_m are non-negative. Although it has not been specifically indicated, it is true that the v_m 's, $E_{k,m}$'s, and $D_m(\mathbf{u})$'s will in general depend on λ .

The practical applicability of the formulae (8)–(12) is limited by the fact that some knowledge of the eigenvalues and eigenvectors is required. The consideration of these eigenvalues and eigenvectors is equivalent to the representation of the 'vector' of the K sea-noise processes at the array elements as the sum of K statistically independent noise 'vectors'. The v_m 's are proportional to the average powers associated with these noise 'vectors'.

3. The Spherically Symmetric Array

Such an array can be loosely described as one for which the number of elements per unit volume is approximately dependent only on the distance of this volume element from the centre of the array and is nearly independent of the direction of the radius vector from the centre to the volume element. As a consequence, the array gain $G_o(\mathbf{u}, \lambda)$ tends to vary only weakly with \mathbf{u} . Hence, for such an array, it is permissible (and convenient) to examine the average $\langle G_o(\mathbf{u}, \lambda) \rangle$ over all unit vectors \mathbf{u} . From eqns. (2), (3) and (4), it is readily seen that

$$\langle G_o(\mathbf{u}, \lambda) \rangle = K \quad \dots\dots(13)$$

when self-noise is absent, i.e. when $\xi(\lambda) = 0$.[†] Thus, the average optimum array gain does not depend on λ nor on the radius, R , of the array. Moreover, it tends to infinity as K increases indefinitely and R/λ remains fixed. This behaviour is in sharp contrast to that of the conventional gain, $G_c(\mathbf{u}, \lambda)$, which tends to a finite limit when R/λ is fixed and K is chosen larger and larger. For example, when the K elements are (approximately) uniformly distributed over the surface of a sphere with radius R , it can be shown that[‡]

$$\lim_{K \rightarrow \infty} G_c(\mathbf{u}, \lambda) = 4b^2 / [\ln(4\gamma b) - \text{Ci}(4b)] \quad \dots\dots(14)$$

where $b = 2\pi R/\lambda$, $\gamma = 1.781\dots$). Hence, in order for $G_c(\mathbf{u}, \lambda)$ to approach infinity, R/λ as well as K must

[†] Formula (13) holds for arbitrary K -element arrays, showing that such arrays can always be so steered that the optimum array gain is at least K when self-noise is absent. Formula (13) remains valid also for non-isotropic noise, provided that in the average $\langle \rangle$ over all unit-vectors weighting proportional to the noise-power from the direction \mathbf{u} be applied. Furthermore, the wavefronts need not be planar. It is merely required that noise wavefronts and signal wavefronts coming from the same direction have identical shapes. This shape is permitted to vary with \mathbf{u} .

[‡] P. Rudnick, unpublished notes. This formula can be derived by noting that $\langle 1/G_c(\mathbf{u}, \lambda) \rangle$ is equal to the average of $q_{k, \lambda}(\lambda)^2$ taken over all K^2 pairs of elements and by replacing this discrete mean by an integration over the spherical surface (cf. eqns. (2) and (5)).

tend to infinity. In particular, $G_c(\mathbf{u}, \lambda)$ is near unity when R/λ is small compared to unity.

When self-noise is present (i.e. when $\xi(\lambda) > 0$), the following approximation holds for large K :

$$G_o(\mathbf{u}, \lambda) \simeq \sum_{n=0}^{\infty} \frac{2n+1}{1 + [\xi(\lambda)/K\mu_n]} \quad \dots\dots(15)$$

where

$$\mu_n = \int_0^R \phi_n^2(2\pi r/\lambda) \rho(r) 4\pi r^2 dr \quad \dots\dots(16)$$

in which $\phi_n(x) = \sqrt{(\pi/2x)} J_{n+1/2}(x)$, and $K\rho(r)$ is the number of array elements per unit volume at distance r from the centre. Formula (15) is obtained when the summations in (3) and (4) are replaced by integrations and when the integral equations resulting from (4) are solved in terms of the functions $\phi_n(2\pi r/\lambda) S_{n,m}(r/r)$, where $S_{n,m}$; $m = 0, 1, \dots, 2n$ denotes the spherical surface harmonics of order n .[§] The same technique leads also to the following approximation

$$\beta_o(v, \mathbf{u}, \lambda) \simeq \left| \frac{1}{G_o(\mathbf{u}, \lambda)} \sum_{n=0}^{\infty} \frac{(2n+1)P_n(\cos \theta)}{1 + [\xi(\lambda)/K\mu_n]} \right|^2 \quad \dots\dots(17)$$

where θ is the angle between the steering direction \mathbf{u} and the signal direction \mathbf{v} , while $P_n(x)$ designates the Legendre polynomial of order n .

Another derivation of (15) can be given based on the formulae (8)–(10). Since $\langle |D_m(\mathbf{u})|^2 \rangle$ is equal to Kv_m (cf. (9) and (2)) it follows that

$$\langle G_o(\mathbf{u}, \lambda) \rangle = \sum_{m=1}^K \frac{1}{1 + [\xi(\lambda)/Kv_m]} \quad \dots\dots(18)$$

This formula holds for any K -element array and is an extension of (13) to the case where self-noise is present. Using the expansion presented for $\exp(2\pi j(\mathbf{u} \cdot \mathbf{r})/\lambda)$ and replacing summations by integrations, it readily follows that the sequence of eigenvalues (v_1, \dots, v_K) of a spherically symmetric K -element array tends to the sequence

$$\mu_0; \mu_1, \mu_1; \mu_2, \mu_2; \mu_2, \mu_2; \mu_2, \mu_2; \mu_3, \dots \quad \dots\dots(19)$$

as K tends to infinity. The eigenvalue μ_n has multiplicity $2n+1$, i.e. μ_n occurs $2n+1$ times in the sequence (19). The corresponding $2n+1$ eigenfunctions are $\phi_n(2\pi r/\lambda) S_{n,m}(r/r)$; $m = 0, 1, \dots, 2n$. Replacing v_m in eqn. (18) by the m th term in the sequence (19) and extending the summation to infinity then yields (15).

The second derivation suggests that it may be more natural not to replace the upper summation limit, K ,

[§] Equation (15) readily follows from eqns. (1–4) and from the following expansion:

$\exp(2\pi j(\mathbf{u} \cdot \mathbf{r})/\lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{2n} j^n \phi_n(2\pi r/\lambda) S_{n,m}(\mathbf{u}) S_{n,m}(r/r)$, where the $S_{n,m}$'s are orthonormal (i.e. $\langle S_{n,m}(\mathbf{u}) S_{n',m'}(\mathbf{u}) \rangle$ is unity when $n = n'$ and $m = m'$ and is zero otherwise (cf. H. Lamb, "Hydrodynamics", p. 512)).

in (18) by infinity but to leave it as it is. This leads to a "terminating" version of (15) whereby the summation in (15) is executed from $n = 0$ to $n = N$ (N being the integral part of \sqrt{K}), with the numerator, $2N + 1$, of the last term replaced by $K - N^2$. Formula (17) can be terminated in the same manner.

In order to compare the performance of the optimum detector with that of the conventional detector, we consider an array whose elements are distributed (approximately) uniformly on the surface of a sphere with radius R . We shall use formula (14) to compute the conventional gain, G_c , and the terminating version of (15) to calculate the optimum gain, G_o . We note that $\mu_n = \phi_n^2(b)$; $b = 2\pi R/\lambda$ in this application (cf. (16)).

For self-noise which is at least 10 dB below the sea-noise, one can expect these two formulae to be accurate to about 0.2 dB or better whenever K is at least 14 and the density $p = \pi K/b^2$ is at least 7 (cf. Appendix 1). Accordingly, we shall limit ourselves to those values of K , p and ξ for which indeed $K \geq 14$, $p \geq 7$ and $\xi \leq 0.1$. To begin with, it follows from eqn. (15) that *the optimum gain tends to infinity whenever K grows to infinity and R/λ remains fixed, even when self-noise is present* (i.e. $\xi(\lambda) > 0$). However, due to the assumed statistical independence of the self-noise processes at the different elements, the optimum gain grows much more slowly with K when self-noise is present than when it is absent. As an example, assuming the self-noise to be 20 dB below the sea-noise ($\xi(\lambda) = 0.01$), Table 1 lists the values of K required to produce an excess of 4 and 6 dB, respectively, of the optimum gain over the conventional gain. The last line in Table 1 provides the values of K needed to achieve the conventional gain given by formula (14). (Consistent with the restraints formulated earlier, K was calculated by requiring p to equal 7.)

Table 1

Number of elements, K , required for 4 dB and 6 dB improvement over conventional gain

$b = 2\pi R/\lambda$	3	6	12	24
Excess = 6 dB	1000	3.2×10^5	1×10^9	1×10^{14}
Excess = 4 dB	50	500	1.5×10^4	1×10^6
Conventional gain	20	80	320	1290

Thus from a practical point of view, it is quite impossible to achieve a gain which is appreciably higher than the conventional gain if R/λ is larger than unity. Stated differently, the conventional detector is practically optimum whenever R/λ exceeds unity.

The use of the optimum instead of the conventional detector can be advantageous only when R/λ is small

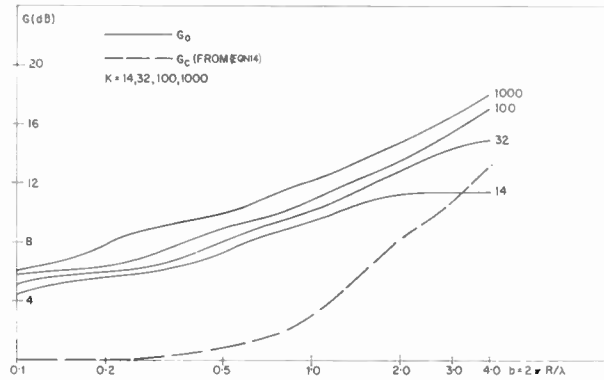


Fig. 1. Gain versus b and K .

(less than unity) because of the rapid decrease of G_c with decreasing values of R/λ in this region. The situation is depicted in Fig. 1. It appears reasonable to select R/λ between 0.08 and 0.16 (i.e. b between 0.5 and 1.0), because then G_o exceeds G_c by about 6 dB for K as low as 14.

This Section will be closed with a few comments on the beam pattern of the optimum detector (eqn. 17). The half-width of its main lobe at the 3 dB-down point is approximately $1.6/\sqrt{G_o}$ radians. For example, when the self-noise is 20 dB below the sea-noise, $K = 32$ and $b = 1$ (i.e. $R/\lambda \approx 0.16$), then $G_o = 10$ and the half-width is about 30 deg. Detailed calculations using eqn. (17) indeed show that the half-width of the main lobe is 30 deg. The main lobe is 18 dB down at 60 deg, while the remainder of the beam pattern (consisting of one side lobe and a back lobe) is down everywhere by at least 17.5 dB. In contrast, the same array has a conventional gain of only 3 dB, and its beam pattern consists of one single lobe which is 3 dB down at 85 deg and only 7 dB down at 180 deg.

4. The Linear K -Element Array

For the special case that *self-noise is absent* ($\xi(\lambda) = 0$), Bryn's method yields results which were already obtained in part by R. L. Pritchard in his study of super-directive beam patterns arising when the output signal-to-noise ratio is maximized.† Pritchard pointed out specifically that the linear array does not tend to behave as a point receiver when the wavelength tends to infinity (or, equivalently, when the array length tends to zero).

Application of Bryn's method leads to the following two formulae which extend some of Pritchard's results:

† R. L. Pritchard, "Directivity of Acoustic Linear Point Arrays", Acoustics Research Laboratory, Harvard University, Cambridge, Mass.; TM 21, 1951. Pritchard obtained formula (20) for the special case of a uniformly spaced linear array with $K = 3$ and steered broadside ($\psi = 90^\circ$).

$$\lim_{\lambda \rightarrow \infty} G_o(\lambda, \psi) = K [P_{K-1}(\cos \psi) P'_K(\cos \psi) - P_K(\cos \psi) P'_{K-1}(\cos \psi)] \dots (20)$$

and

$$\lim_{\lambda \rightarrow \infty} \beta_o(\lambda, \theta, \psi) = \left\{ \frac{K [P_{K-1}(\cos \psi) P_K(\cos \theta) - P_K(\cos \psi) P_{K-1}(\cos \theta)]}{\lim_{\lambda \rightarrow \infty} G_o(\lambda, \psi) [\cos \theta - \cos \psi]} \right\}^2 \dots (21)$$

where ψ is the angle between the array and the direction in which it is steered and where θ is the angle between the array and the direction of the received signal. $P_n(x)$ and $P'_n(x)$ denote the Legendre polynomial of degree n and its derivative. It is worthwhile pointing out that eqns. (20) and (21) hold for uniform as well as arbitrary non-uniform spacing of the K elements.

To derive (20) and (21) we note that

$$q_{k,h} = \sum_{n=0}^{\infty} (2n+1) \phi_n(a_k) \phi_n(a_h)$$

where a_k is the product of $2\pi/\lambda$ and the co-ordinate of the k th element and where ϕ_n is as in Section 3. With Z_k as in eqn. (4) we define

$$\zeta_n = \sum_{k=1}^K Z_k \phi_n(a_k)$$

It follows that ζ_n is a linear combination of $\zeta_0, \dots, \zeta_{K-1}$, whose coefficients tend to zero as λ approaches infinity, whenever $n \geq K$. Using the expansion

$$p_h^* = \sum_{n=0}^{\infty} (-j)^n (2n+1) \phi_n(a_h) P_n(\cos \psi)$$

it is seen from eqn. (4) that, in the limit as $\lambda \rightarrow \infty$, ζ_n will equal $(-j)^n P_n(\cos \psi)$ when $n < K$ and zero when $n \geq K$, provided that $\xi(\lambda) = 0$. Substitution of this result in eqns. (3) and (6) yields (20) and (21).

We shall examine the cases of broadside steering ($\psi = \pi/2$) and of endfire steering ($\psi = 0$) in some detail. For *broadside* steering it follows from eqn. (20) that the limiting gain is

$$[P'_{2N+1}(0)]^2 (\simeq (2/\pi)(2N+1.5))$$

where $2N+1$ is the largest odd integer not exceeding K . According to eqn. (21) the limiting beam pattern is

$$[P_{2N+1}(\cos \theta) / P'_{2N+1}(0) \cos \theta]^2$$

Thus, in the limit, the array indeed does not behave as a point receiver. The beam pattern is a surface of revolution with the array as its axis. In addition to the main lobe, it has $2N$ side lobes arranged symmetrically about the main lobe. Near the main lobe, the beam pattern is approximated by $(\sin x/x)^2$, where

$$x = [2N+1.5] [(\pi/2) - \theta].$$

Thus, the semi-width of the main lobe at the 3 dB-down point is $1.4/(2N+1.5)$ radians. Also, we see that the maximum response on the two side lobes contiguous to the main lobe is 13.5 dB below that on the main lobe.

For *endfire* steering it follows from eqn. (20) that the limiting gain is K^2 . This is substantially larger than for broadside steering.† It is seen from (21) that the limiting beam pattern is identical to the mean beam pattern of a spherically symmetric K^2 -element array (apply the "terminating" version of eqn. (17), choosing $\xi(\lambda)$ zero). In addition to the main lobe and the back lobe, the beam pattern has $K-2$ side lobes. Near the main lobe the beam pattern is approximated by $[2J_1(x)/x]^2$, where $x = (2K+1)\sin(\theta/2)$. Hence the semi-width of the main lobe at the 3 dB-down point is $1.6/(K+0.5)$ radians, while the maximum response on the first side lobe is 17.5 dB below that on the main lobe.

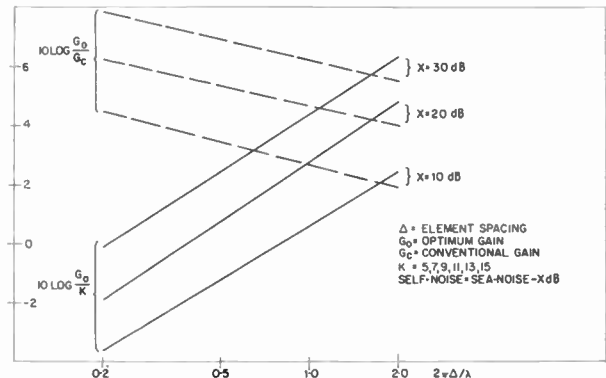


Fig. 2. K -element endfire array.

Since in the absence of self-noise, the endfire array is superior to the broadside array both insofar as array gain and suppression of the first side lobe are concerned, it is worthwhile to examine the performance of the endfire array *in the presence of self-noise*. Assuming uniform spacing, G_o was calculated for the following values of the pertinent parameters: $K = 5, 7, 9, 11, 13, 15$; $\xi = 0.001, 0.01, 0.1$ (corresponding to self-noise respectively 30, 20, 10 dB below sea-noise); $s = 0.2, 0.5, 0.7, 1.0, 1.3, 1.5, 2.0, 2.5, 3.0, 3.5$, where s is equal to $2\pi/\lambda$ times the distance between successive elements. The results are depicted in Fig. 2. It is seen that G_o/K and G_o/G_c are nearly independent of K and depend largely on ξ and s only. As could be expected, the presence of self-noise substantially degrades the performance. Nevertheless, the optimum

† We conjecture that the optimum gain of any K -element array cannot exceed K^2 in the presence of isotropic sea-noise and self-noise.

detector can appreciably outperform the conventional detector. For example, when the self-noise is 20 dB below the sea noise ($\xi = 0.01$), and $K = 7$, $s = 0.5$ (making a total array length slightly less than 0.5λ) the conventional gain is 3.9 dB while the optimum gain is 9.3 dB, (without self-noise the optimum gain would have been 16.8 dB).

In order to examine the beam pattern, these patterns were calculated for $K = 7, 11, 15$; $\xi = 0.1, 0.01$; and $s = 0.2, 0.5, 1, 2, 3$. For $s \leq 2$ the semi-width of the main lobe at the 3 dB-down point was approximated by $1.6/\sqrt{G_o}$ radians with an accuracy of one part in ten or better, while the maximum response on the first side lobe was between 15.4 and 17 dB below that on the main lobe (except for $K = 7$, $s = 0.2$, $\xi = 0.1$, where the first side lobe coincided with the back lobe and the maximum response was down by only 8.5 dB). We note that the above approximation to the semi-width is also valid for the case considered earlier where self-noise is absent ($\xi = 0$) and $\lambda \rightarrow \infty$ (or: $s \rightarrow 0$). The number of side lobes when self-noise is present appears to be considerably less than when self-noise is absent. For the self-noise considered here (10 to 20 dB below sea-noise) the side lobe structure seems to be largely determined by the total array length. As an example, when $K = 7$, $s = 0.5$ and the self-noise is 20 dB below the sea-noise, the optimum beam pattern has one main lobe, one back lobe and one side lobe. The main lobe is down 3 dB when $\theta = 32$ deg and 15.9 dB when $\theta = 60$ deg. Thereafter, the beam pattern is down everywhere at least 15.9 dB. In contrast, the conventional beam pattern of the same array has a main lobe and a back lobe, but no side lobe. Its main lobe is down 3 dB at $\theta = 80$ deg, and at $\theta = 123$ deg it is down 15.9 dB. Thereafter the beam pattern remains at least 15.9 dB down.

The preceding results show that the use of the optimum detector instead of the conventional detector when K and the total array length are small, may result in superior array performance.

5. The Ring of K Equally-Spaced Elements

This array is of interest because the eigenvectors and eigenvalues of its $q_{k,h}$ -matrix can be described explicitly. They are as follows:

$$E_{k,m} = \frac{1}{\sqrt{K}} \exp(2\pi jkm/K);$$

$$k = 1, \dots, K; \quad m = 0, 1, \dots, K-1 \quad \dots\dots(22)$$

and

$$v_m = \frac{1}{2b} \int_0^{2b} \sum_{p=-\infty}^{\infty} J_{2m+2pK}(t) dt; \quad m = 0, 1, \dots, K-1$$

$$\dots\dots(23)$$

where J_q denotes the Bessel function of order q and where $b = 2\pi R/\lambda$, R being the ring's radius.

Also

$$D_m(u) = \sqrt{K} \sum_{p=-\infty}^{\infty} J_{m+pK}(b \cos \theta) (j \exp(-j\phi))^{m+pK}$$

$$\dots\dots(24)$$

where θ is the angle between the unit vector u and the plane of the ring while ϕ is the angle between the radius vector to one of the elements and the projection of u on the plane of the ring. Optimum and conventional array gains and beam patterns can be calculated by substituting (23) and (24) in (10), (11) and (12). We shall briefly discuss the results which can be obtained in this manner. The derivations are straight-forward and will be omitted.

We begin with the case where the steering direction lies in the plane of the ring ($\theta = 0$). As in the case of the spherically symmetrical array, the optimum gain tends to infinity as K increases indefinitely and R/λ remains fixed, even when self-noise is present. This result can be shown in a mathematically rigorous manner. However, as before, the growth of the gain with K is exceedingly slow in the presence of self-noise. For example, when $b = 10$ (i.e. $R/\lambda \simeq 1.6$), and the self-noise is 20 dB below the sea-noise ($\xi = 0.01$), about fifty elements suffice to attain a conventional gain of 13 dB, but three million elements are needed to achieve an optimum gain of 19 dB.

When self-noise is absent and R/λ is small compared to unity, the optimum gain is approximately $0.53(K+1)^{3/2}$, showing that the ring is intermediate between the spherically symmetric array ($G_o = K$) and the endfire linear array ($G_o \rightarrow K^2$ when $\lambda \rightarrow \infty$).

Next we consider the case where the steering direction lies outside the plane of the ring ($\theta \neq 0$). In that case the optimum gain remains finite as K tends to infinity. In particular, the optimum and the conventional detectors coincide when the steering direction is perpendicular to the plane of the ring. The optimum gain of an infinite-element ring tends to $-1 + 2\text{cosec}^3 \theta$ as R/λ approaches zero. Also, an expression for the limiting beam pattern can be derived. When this expression is used to calculate the array response in the direction obtained by projecting the steering direction onto the plane of the ring, one finds that the response in the projected direction is about 9 dB higher than in the steering direction whenever θ is small ($|\theta| < 30^\circ$). This unusual feature is a consequence of the rapid increase of the limiting gain as θ approaches zero, and of the fact that the tangent planes to the beam pattern and the 'gain' surface are parallel at the respective points of intersection with the steering direction (the 'gain' surface being obtained by plotting G_o as a function of the steering direction (θ, ϕ)).

6. Conclusions

For the 'normal' conditions postulated in Section 1, little practical benefit appears to be gained from the use of the optimum detector when self-noise is present and the maximum array dimension is of the order of one wavelength or more.

The optimum detector seems advantageous only when the array dimensions are small compared to the wavelength and the number of elements is moderate (about twenty, say).

Even in this exceptional case the margin of preference of the optimum detector over the conventional detector is decreased further as a result of the fact that the values of the cross-spectral coefficients, $q_{k,h}$, used in the design of the optimum detector will generally be different from the true values, due to measurement errors and variability of the medium. Instrumentation imperfections will cause additional performance degradation.

7. Acknowledgments

The author is indebted to Dr. Philip Rudnick, University of California, San Diego, Marine Physical Laboratory of the Scripps Institution of Oceanography, San Diego, California, and to Mr. F. Bryn of the Norwegian Defence Research Establishment, K.J.V., Horten, Norway, for suggesting the topic discussed herein and for several helpful conversations.

The paper represents research partially supported by the Office of Naval Research under contract Nonr 2216(05).

8. Appendix:

Accuracy of Formulae (14) and (15)

This appendix is devoted to a discussion of the accuracy of eqn. (14) and of the terminating version of eqn. (15) when applied to an array whose elements are (approximately) uniformly distributed on the surface of a sphere with radius R .

The accuracy of the terminating version of (15) is determined by the closeness with which the first K terms of the sequence (19) approximate the eigenvalues v_1, \dots, v_K . In order to gain some insight into this matter it is instructive to examine the special array consisting of fourteen elements so arranged on the surface of a sphere of radius R , that eight of them are the vertices of a cube while the other six are the extremities of the three diameters parallel to the edges of the cube. The fourteen eigenvalues v_1, \dots, v_{14} can be readily calculated in terms of the parameter $b = 2\pi R/\lambda$ and can then be compared with the first fourteen terms of the sequence (19). Since the fourteen elements are located on the surface of a sphere of radius R , μ_n must be obtained from eqn. (16) with $4\pi r^2 \rho(r)$ being the δ -function having its infinite peak at $r = R$, i.e. $\mu_n = \phi_n(b)^2$. Table 2 provides the comparison of the v 's and μ 's. As could be expected, the agreement is good for small values of b and worsens as b increases. Also, for fixed b , the agreement deteriorates as one progresses from the first to the fourteenth term of the sequence (19). Nevertheless, the agreement is rather satisfactory for values of b as high as 3. Consequently, one can anticipate good correlation between the actual value of the averaged optimum array gain and its

Table 2
Comparison of the v 's and μ 's.

$b = 0.2$		$b = 2.0$		$b = 3.0$	
v	μ	v	μ	v	μ
0.986738	0.986738	0.206708	0.206705	0.002259	0.002213
0.004409	0.004409	0.189602	0.189571	0.120090	0.119493
0.004409	0.004409	0.189602	0.189571	0.120090	0.119493
0.004409	0.004409	0.189602	0.189571	0.120090	0.119493
0.7576×10^{-5}	0.7071×10^{-5}	0.042354	0.039382	0.098127	0.089184
0.7576×10^{-5}	0.7071×10^{-5}	0.042354	0.039382	0.098127	0.089184
0.6734×10^{-5}	0.7071×10^{-5}	0.037758	0.039382	0.088965	0.089184
0.6734×10^{-5}	0.7071×10^{-5}	0.037758	0.039382	0.088965	0.089184
0.6734×10^{-5}	0.7071×10^{-5}	0.037758	0.039382	0.088965	0.089184
1.2843×10^{-8}	0.5779×10^{-8}	0.008194	0.003687	0.051379	0.023120
0.9173×10^{-8}	0.5779×10^{-8}	0.005867	0.003687	0.037212	0.023120
0.9173×10^{-8}	0.5779×10^{-8}	0.005867	0.003687	0.037212	0.023120
0.9173×10^{-8}	0.5779×10^{-8}	0.005867	0.003687	0.037212	0.023120
0.1024×10^{-10}	0.5779×10^{-8}	0.000708	0.003687	0.011309	0.023120

approximate value calculated from the 'terminating' version of eqn. (15). In fact, both values agreed to within 0.2 dB for b ranging up to 3.5 and for self-noise ranging from 10 dB to 30 dB below the sea-noise (i.e. for ξ from 0.1 to 0.001).

In order to generalize these results to other values of K and b , we note that the accuracy of the 'terminating' version of (15), when applied to K elements (approximately) uniformly distributed on the surface of a sphere of radius R , is controlled principally by the average number of elements per unit area, with the wavelength λ taken as unit length. This density, p , is given by $\pi K/b^2$. When K is 14 and b is 3, p is nearly 5. Hence, from the previous results it is reasonable to infer that the accuracy of the "terminating" version of (15) is on the order of 0.2 dB whenever K is at least 14, p is at least 5, and the self-noise is between 10 and 30 dB below the sea-noise.†

The conventional gain of this K -element array can be calculated from the formula (14). Again, this

† The parameter p also controls the accuracy of (15) when the elements are distributed in the volume of a sphere of radius R . Of course its geometric meaning is in this case slightly different from the one above.

formula is an approximation for which the accuracy depends largely on p . The accuracy was examined for $K = 14$ by comparing the value of G_c obtained from (14) with the reciprocal of the exact average value of $1/G_c$ calculated for the afore-mentioned special 14-element array by means of the following formula:‡

$$\langle 1/G_c(u, \lambda) \rangle = \sum_{m=1}^K v_m^2 \quad \dots\dots(25)$$

Agreement to within 0.1 dB was found to exist for values of b as large as 2.5. Beyond that value the exact gain dropped noticeably below the value given by (14). Since $p = 7$ when $K = 14$ and $b = 2.5$, it is again reasonable to expect similar accuracy in (14) whenever $K \geq 14$ and $p \geq 7$.

‡ Self-noise has been ignored because its effect on G_c is quite minor. In order to account for self-noise, the term $\xi(\lambda)/K$ must be added to the right-hand side of (25). Formula (25) follows from (11) and from $\langle |D_m(u)|^2 \rangle = K v_m$.

Manuscript first received by the Institution on 12th June 1962 and in final form on 8th December 1962. (Paper No. 851/SS20.)

© The British Institution of Radio Engineers, 1963

The Effect of Ice on Long Range Underwater Sound Propagation

By

J. D. MACPHERSON, Ph.D.†

Presented at the Symposium on "Sonar Systems" in Birmingham on 9th-11th July, 1962.

Summary: A study has been made of the effect of ice cover on long range (100 n miles) acoustic propagation at low frequencies in an area having an approximate constant water depth of 30 fathoms. Explosive charges weighing 1.8 lb dropped from an aircraft were used as acoustic sources and the acoustic pulses were detected on a bottomed barium titanate hydrophone. Comparative measurements have been made with identical water velocity structures with no ice and with a variety of different ice conditions. Large variations of attenuation have been observed depending upon the type of ice prevalent in the area.

1. Introduction

During the winters of 1960-1 and 1961-2 a series of experiments were carried out off the north coast of Prince Edward Island to measure the effect of ice cover on acoustic propagation. This area was selected for ease of access and because it is fairly typical of coastal water which is subjected to seasonal ice. The water depth in this area is about 30 fathoms and the bottom consists of a medium velocity rock with little or no unconsolidated sediment. Ice usually begins to form in mid-January and finally vanishes in May or June.

With a typical winter air temperature in the Gulf of St. Lawrence of 0° F, the initial directly frozen ice sheet seldom exceeds a foot in thickness. As the ice is slowly moved about by wind and water currents, the edges of the floes ride up on each other and freeze in position. This is called rafting and is the quickest and commonest way of forming thick winter ice. The first major storm after the ice has formed causes the rafted ice to buckle up and again refreeze. The ice then becomes ridged winter ice and depending on the conditions the thickness may then reach thirty feet. This type of ice is very rough both on its top and bottom surfaces. During April and May there is considerable melting but the ice generally remains until it is broken up in a storm and carried out into the Atlantic.

2. Experimental Method

Propagation experiments were carried out at approximately monthly intervals throughout the ice season so as to obtain measurements under a variety of environmental conditions.

Figure 1 shows the location of the barium titanate hydrophone which was mounted a few feet above the

bottom in about 18 fathoms of water. Beyond this position the water depth increases gradually over the next 20 nm until it reaches 30 fathoms. The hydrophone was directly connected by a submarine cable to a hut on the shore where the signals received were amplified and recorded on paper and magnetic tape recorders.

In each propagation experiment an aircraft flew approximately along the bearings shown in Fig. 1 and dropped pressure-actuated 1.8 lb TNT charges set to fire at 50 ft into holes in the ice. The number of charges dropped in any one experiment varied depending on the number of holes available in the ice but in general about 100 charges were fired during the four runs. The charge locations and thus the ranges from the hydrophone were determined from Decca readings taken in the aircraft as the charges were released. Out to 30 nm the range was also calculated from the time difference between the water and bottom or

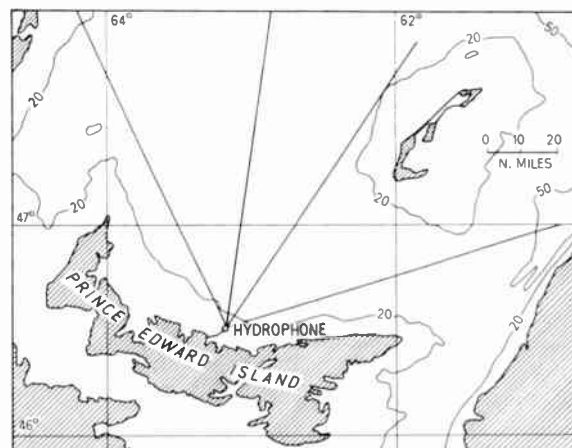


Fig. 1. Map of experimental area showing hydrophone location and approximate bearings of the propagation runs.

† Naval Research Establishment, Dartmouth, N.S., Canada.

ground arrivals using the previously measured bottom velocity¹ and the known water velocity. In general, the differences between the ranges measured by this technique and those obtained from the Decca readings did not exceed $\frac{1}{2}$ nm.

The acoustic pulses recorded on tape were played back through seven octave filters covering the frequency range of 25–3200 c/s. The outputs of the filters were then squared, integrated and converted to decibels relative to 1 erg/cm² for a 1 c/s band by means of the known hydrophone sensitivity and the acoustic impedance of the medium.

3. Theory

In shallow water acoustic energy is propagated to long ranges by a series of reflections at the surface and the bottom. At frequencies below 5 kc/s the effect of direct absorption of acoustic energy by sea water is negligible compared to these reflection losses. If the surface and bottom are perfect parallel reflectors then the acoustic energy from a point source will spread out cylindrically. Thus, if the loss in excess of cylindrical spreading is determined from the measured energy level by subtracting the idealized cylindrical spreading, a quantity is obtained which is related to the reflection losses occurring at the surface and the bottom. A typical plot of level against

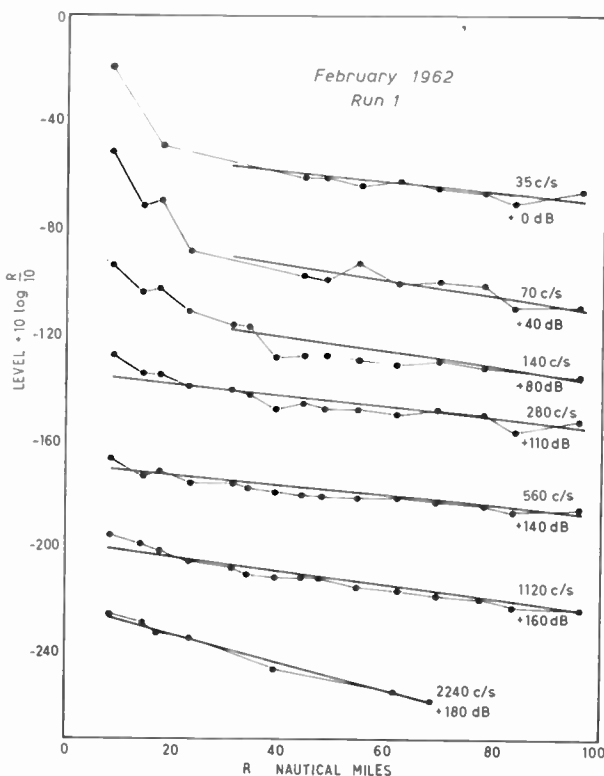


Fig. 2. A plot of corrected level against range for measurements made in February 1962 under young smooth ice.

range corrected for cylindrical spreading is shown in Fig. 2 where each octave band is labelled with its geometric mean frequency and is displaced in the vertical by an amount indicated in the adjacent column.

The straight lines which have been fitted by eye to these plots show that to a good approximation the energy loss per nm is constant with range beyond about 30 nm. The slope of this line is defined as the attenuation coefficient.

The relation between the attenuation coefficient and the surface and bottom reflection losses can be obtained by considering the propagation mechanisms. Bathythermographs taken by helicopters showed that the water remained isothermal throughout all the experiments. Thus the acoustic velocity increases slightly with depth, because of the hydrostatic pressure effect, forming a surface duct. When the wavelength of the acoustic energy is very small compared to the water depth, propagation to long ranges will be controlled by this surface duct as bottom-reflected energy will be rapidly attenuated. Under these conditions propagation can be approximately described by a narrow bundle of rays which is refracted so as to just miss the bottom. This energy will suffer a fixed number of reflections per nm at the surface independent of frequency. The angle at which these reflections occur and the number of reflections per mile can be calculated from simple ray theory. This model of the propagation will not be applicable below the cut-off frequency of the surface duct which was estimated² to be about 400 c/s.

When the wavelength of the acoustic energy is of the same order of magnitude as the water depth, the small velocity gradient in the water can be neglected and the medium treated as an isovelocity liquid overlying a semi-infinite solid. Thus propagation to long ranges can be represented as the sum of a number of normal modes. The phase and group velocities of these modes were calculated as a function of frequency for the known bottom compressional wave velocity of 9 860 ft/s and for several shear wave velocities. The calculated group velocities were then compared with measured values of the group velocity as a function of frequency obtained by playing back about one hundred selected shots through a Kay Vibralyzer. This instrument presents a plot of frequency against time with the acoustic intensity represented by the intensity of the mark on the paper. It appeared from the comparison that a low shear wave velocity of about 5 240 ft/s gave the closest agreement. An average velocity of 4 720 ft/s was assumed for the water. Thus using this model of the medium and the plane wave representation of a normal mode an average grazing angle and number of reflections per mile were computed for each octave up to 800 c/s. The number of reflections per mile

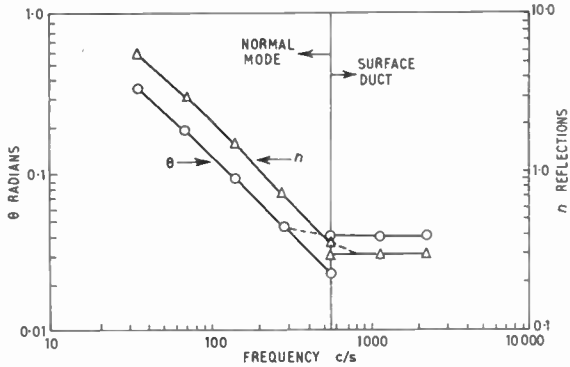


Fig. 3. The number of reflections per nautical mile and the grazing angle associated with the geometric centre frequency of each octave band.

and the grazing angles as calculated by ray and normal mode theory are shown in Fig. 3. The two methods of calculation overlap in the 400–800 c/s octave and in this intermediate region values were arbitrarily selected as shown by the dotted lines so as to produce a smooth transition. The quantities shown in this figure were used to convert the experimentally determined attenuation coefficients into a measure of the average loss per reflection.

4. Experimental Results

Experiments were carried out on 12th January 1961 and on 11th January 1962 with open water over the whole area of interest. The sea states during the two experiments were quite similar and the average wave height was estimated as 2–4 ft. The results from the two experiments and from the four different bearings were not significantly different and all the data were corrected for cylindrical spreading and combined to make a composite plot of level against range. A least-squares technique was used to fit straight lines to these plots and a spread or error associated with these straight line fits was calculated from the ratio of $(1/N \sum (L - L_0)^2)^{1/2}$ to the range interval over

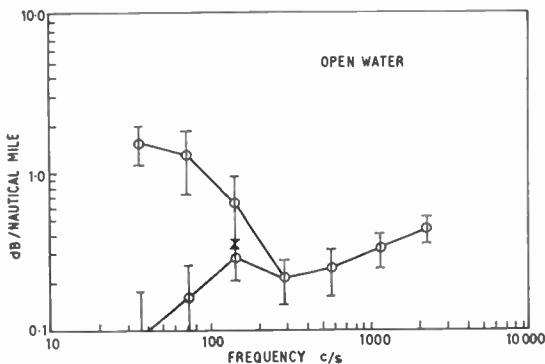


Fig. 4. Attenuation coefficients measured in January under open water conditions.

which the line was fitted. (L is the measured level + $10 \log_{10} (R/10)$ and L_0 is the level predicted by the straight line at the same range.) The attenuation coefficients determined in this manner together with their appropriate ‘errors’ are shown in Fig. 4. The double values at low frequencies arise from the different attenuation coefficients measured at different ranges from the hydrophone. At short ranges the higher order modes predominate giving large values of the attenuation coefficients. Beyond 30 nm the first mode is the only one of importance and gives the lower values shown in this figure. At higher frequencies this effect was not observed over the range interval of 10–100 nm studied in these experiments.

The high-frequency surface duct controlled attenuation coefficients were used together with the number of reflections per nm to calculate an average loss per reflection at the sea surface. Figure 5 shows the experimental loss in decibels per reflection plotted against the product of frequency and average wave

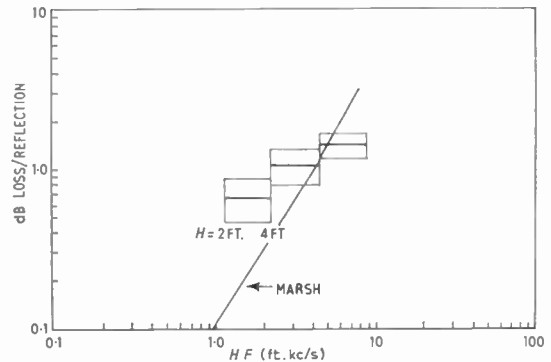


Fig. 5. The loss in dB per reflection plotted against the product of frequency and average wave height for open water conditions.

height together with a theoretically predicted curve by Marsh, Schulkin and Kneale³ based on a scattering mechanism. The estimated spread of wave heights is indicated by the width of the symbol used for each octave.

Experiments were also carried out on 7th February 1961 and 5th February 1962 both with relatively smooth rafted ice about one foot thick over the whole area. As in the open water experiments there was little difference among any of the runs and the results were combined to give the attenuation coefficients as shown in Fig. 6. These attenuation coefficients are on average about 1/10 dB/nm greater than those measured under open water but the most outstanding feature is the relatively small effect of smooth ice on the propagation. This is in agreement with results obtained by A. R. Milne in Barrow Strait under smooth seasonal ice about seven feet thick.⁴

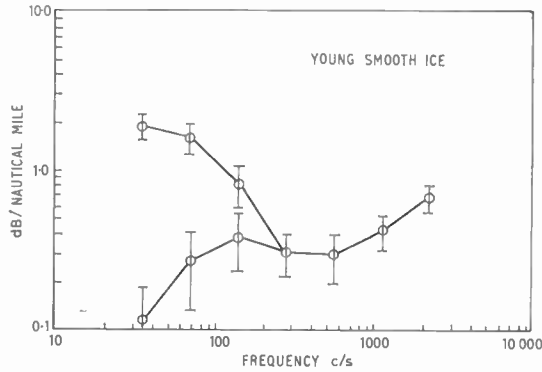


Fig. 6. Attenuation coefficients measured under young smooth ice conditions.

Three experiments were carried out under rough ice conditions on 7th April 1961, 27th April 1961 and 14th March 1962. The results of these experiments were combined to give the attenuation coefficients as shown in Fig. 7. The higher losses encountered under these conditions account for the absence of double values of the attenuation coefficient at low frequencies as the range interval over which the higher angle modes are important is reduced under these conditions to ranges of less than 10 nm.

At high frequencies the surface reflection loss is the only important mechanism contributing to the attenuation coefficient and thus as before the number of reflections per mile can be used to convert the attenuation coefficient to an average loss per reflection. At low frequencies it is necessary to consider the energy lost into the bottom. This loss has already been approximately measured by the experiments made in open water and thus the additional loss due to the rough ice surface can be calculated by subtracting the attenuation coefficient measured under open water conditions from the attenuation coefficient measured under rough ice. Although this is not a

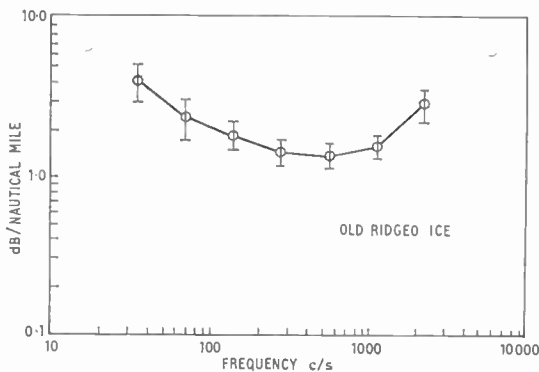


Fig. 7. Attenuation coefficients measured under old ridged ice conditions.

very precise correction, the magnitude of this correction is small as the attenuations measured under open water are considerably less than those measured under rough ice. This corrected attenuation coefficient can now be converted to an average loss per reflection at the ice surface by means of the previously computed number of reflections per nm.

At low frequencies the grazing angle is a function of frequency and thus the reflection losses are occurring at a variety of grazing angles depending upon the frequency. If it is assumed that the loss per reflection is directly proportional to the grazing angle for small grazing angles then the calculated losses can be normalized to the same angle for all frequencies. Figure 8 shows the loss in decibels per reflection at the rough ice surface normalized to an angle of 0.038 radians which is the grazing angle for the high-frequency surface duct energy. This reflection loss is plotted against the product of frequency and 'average ice roughness'. This 'roughness' has been set arbitrarily at 20 ft which appeared to be reasonable for the type of

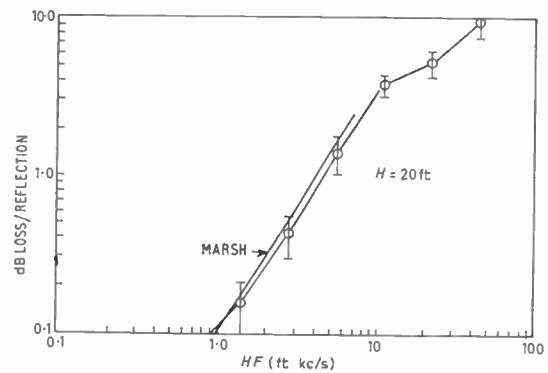


Fig. 8. The loss in dB per reflection at the old ridged ice surface.

ice present during the experiments. The heavy solid line is a theoretical curve predicted by Marsh, Schulkin and Kneale³ for specular reflection losses at a rough sea surface and is valid for losses up to 3 dB per reflection. The apparent general agreement between the measured losses and the theoretical curve indicates that the reflection losses are probably caused by scattering at the rough ice surface rather than by absorption in the ice.

One experiment was carried out when the ice was partially broken up on 12th April 1962. Attenuation coefficients measured in this experiment were intermediate in magnitude between those measured under rough ice and open water which is consistent with the previous results. No attempt was made to interpret these results as average losses per reflection, as the nature of the ice was too varied to warrant this approach.

5. References

1. J. D. Macpherson, "Seismic refraction measurements of the bottom structure off the north shore of Prince Edward Island", *Bull. Seis. Soc. Amer.*, 52, p. 399, 1962.
2. C. L. Pekeris, "Theory of Propagation of Explosive Sound in Shallow Water". Geological Society of America Memoir 27, 1962.
3. H. W. Marsh, M. Schulkin and S. G. Kneale, "Scattering of underwater sound by the sea surface", *J. Acoust. Soc. Amer.*, 33, p. 334, 1961.
4. A. R. Milne, "Shallow water under ice acoustics in Barrow Strait", *J. Acoust. Soc. Amer.*, 32, p. 1007, 1960.

Manuscript received by the Institution on 27th June 1962. (Paper No. 852/SS7.)

© The British Institution of Radio Engineers, 1963

POINTS FROM THE DISCUSSION

Dr. S. S. Srivastava: The increase in attenuation in the old ridged ice surface is due to scattering of the angle surface of ice. Has any attempt been made to study this roughness and correlate it with the increase in attenuation?

The author (in reply): Aerial photographs were taken of the upper surface of the ice and these were used to classify the ice into several broad categories. No measurements of the topography of the lower ice surface have been made.

Mr. M. Schulkin: It has been found in work under the Arctic ice with 50 to 200 ft underside peaks that the same laws of scattering behaviour are observed as are indicated in this paper.

The author (in reply): The work described in this paper is consistent with measurements made in the Arctic.

Mr. D. E. Weston: Dr. Macpherson pointed out that at low frequencies the attenuation was greater at close ranges than at long ranges. Let us concentrate on the 35 c/s curve of Fig. 2. If the fitted straight line is extrapolated back it gives a measure of the excitation of the first mode. Measured close range values lie some 30 dB above the extrapolated line, explicable if 1000 modes were permitted. However, there are only about five permitted propagating modes. No suggested explanations for the discrepancy are offered here, but Dr. Macpherson may care to comment. Discrepancies of this sort have been noted elsewhere, and it is thought likely that there are several contributing causes.

The author (in reply): At ranges of 5–30 nm the measured attenuation coefficient at low frequencies is several times larger than that measured from 30–100 nm (see Figs. 2 and 4). This larger attenuation coefficient is consistent with the greater number of reflections per nm and the higher grazing angles associated with the higher of the four or five permitted modes of propagation. If this curve is extrapolated back to 1 nm the resulting levels are quite

consistent with cylindrical spreading from 100 yards with a small (< 5 dB) additional attenuation loss.

Mr. M. J. Daintith: The Schulkin and Marsh curve for surface loss per reflection applies to deep water, i.e. to a condition in which energy scattered out of the channel is completely lost. There should be a good correlation with the under-ice measurements in shallow water, therefore, only if the bottom is perfectly absorbing. Observed results should therefore be corrected by subtracting from the total energy that which is reflected from the bottom, not by adding that which is absorbed.

Should not the curves in Fig. 2 merge from a 10 log *R* to a 20 log *R* law at short ranges?

The author (in reply): The method used in discussing these results uses an ideal model of two perfect infinite reflectors as the surface and the bottom. Thus compared to this ideal model energy is lost on reflection at the surface and the bottom. Non-specular forward scattered energy is automatically included in the measurements although it is believed that this contribution is small compared to the specular component.

At ranges of a few water depths, the curves would presumably merge from 10 log *R* to 20 log *R*.

Mr. D. Davies: Did the fact that the hydrophone was situated over a sloping bottom affect the records in any way? For instance, would there be any reflection of the propagated modes from points between the hydrophone and the shore?

The author (in reply): Studies of the dispersion of shots from different ranges indicate that the various modes accommodate themselves to the gradual changes in water depth. The fraction of energy reflected in this process is very small compared to the amount of energy lost by other means.

Determination of Sulphur Content of Hydrocarbons by Bremsstrahlung Absorption measurement

By
T. B. ROWLEY†

Presented at the Convention on "Electronics and Productivity" in Southampton on 17th April 1963.

Summary: The paper discusses the usefulness of β -excited sources of electro-magnetic radiation (bremsstrahlung) in a radiation absorption measuring system, designed primarily to indicate continuously the sulphur content of refinery hydrocarbon streams. A conventional Geiger counter/ratemeter system is used to give simplicity of design and reliability of operation of the detector/indicator equipment. An accuracy of better than $\pm 0.02\%$ weight sulphur over long periods has been obtained under refinery operating conditions. A second equipment, based on the same principle of operation, but designed for laboratory use on a wide range of base stocks, is also described, where sulphur determination is independent of density, and hydrogen/carbon ratio error effects are much reduced.

1. Introduction

Gauges using the radiation absorption characteristics of gamma and bremsstrahlung electro-magnetic sources for density and thickness measurement have been widely reported^{1, 2, 3} and are in common use throughout industry. This type of gauge comprises, basically, a radiation source and detector placed on opposite sides of the material to be measured, and a radiation intensity indicator (Fig. 1). Changes in mass of the measured material

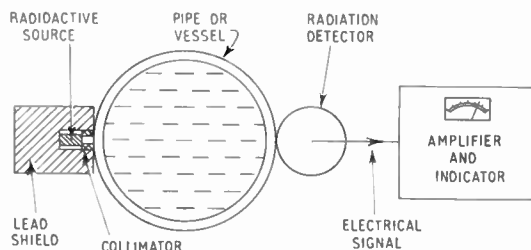


Fig. 1. Schematic diagram of gamma density gauge.

give rise to changes in detected radiation intensity and thus allow an indication of thickness or density of the material under given conditions; in the case of a density gauge using cobalt 60 or caesium 137 radiation, the minimum detectable density change will be of the order of 0.2% of mean density under optimum conditions. Consideration of the relationship of mass absorption co-efficient to radiation energy and atomic number of the absorber (Fig. 2) shows that gauges using Co^{60} and Cs^{137} are sensibly independent of variation in atomic number of the

measured material over a very wide range. However, at a radiation energy of less than 100 keV it will be seen that the mass absorption co-efficient rises very rapidly with increase of atomic number and at 10 keV, for example, the absorption co-efficient of sulphur is approximately 10 times that of carbon and 85 times that of hydrogen.

An equipment using a radiation source of photons of mean energy around 10 keV would show far greater sensitivity to changes in the sulphur content

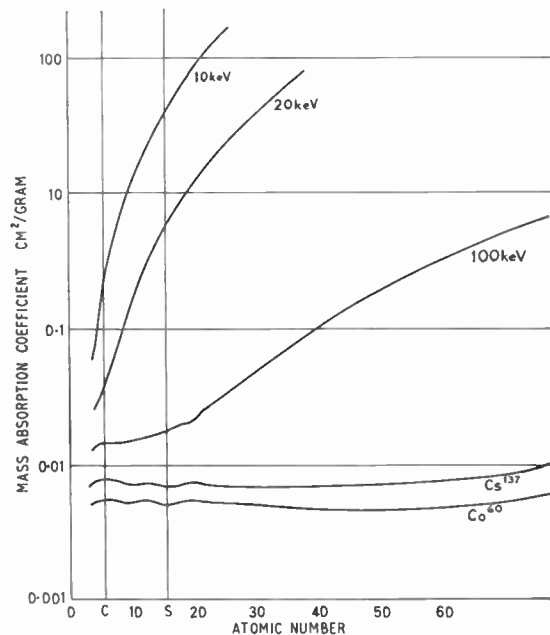


Fig. 2. Mass absorption against atomic number curve for Co^{60} , Cs^{137} , 100, 20 and 10 keV.

† Isotope Developments Ltd., Beenham, Reading, Berkshire.

of a hydrocarbon material than would be expected to arise from the radiation absorption change due to straightforward total mass change of the material. This paper discusses the parameters affecting the design and choice of components for this type of equipment.

2. Operational Requirements of the Equipment

To meet the requirements of refinery plant operation, the following essential design points were specified:

(1) The detector and any associated circuits should be 'intrinsically safe' from an electrical point of view, or should be suitable for housing in certified flame-proof enclosures.

(2) The radiation source should be of long half-life to avoid the need for frequent replacement of the source or for recalibration of the equipment.

(3) The equipment should be capable of making measurements on streams in which the pressure is up to 200 lb/in².

(4) The equipment should be capable of operation by unskilled personnel whilst measuring sulphur contents from 0.05% to 4% weight with a facility for showing smaller ranges (e.g. 0.05% to 0.5%) for particular applications.

(5) The equipment should operate satisfactorily with the detector positioned up to 300 feet from the main amplifier/indicator.

(6) For the plant applications considered, the base stock would be of virtually constant, or known density and carbon/hydrogen ratio. In laboratory applications however, a wide range of densities and carbon/hydrogen ratios would be encountered, and therefore sulphur determination in this equipment should be sensibly independent of both density and carbon/hydrogen ratio of the sample.

3. General Design of the Equipment

The equipment designed to meet the plant requirements comprises a tritium-zirconium source, a flow

cell, an x-ray sensitive Geiger-Müller tube, a quench amplifier and an industrial ratemeter indicator with a suitable recorder output, as shown in Fig. 3.

The radiation source, in the form of a flat disc, is housed adjacent to one 'window' of the flow cell, so that radiation will penetrate through the fixed 'thickness' of the hydrocarbon stream into the Geiger-Müller tube. Impulses from the G-M tube are defined and amplified by the quench amplifier housed close to the detector and passed to the ratemeter; this unit is provided with backing-off and scale expansion facilities so that any given count rate may be set to give zero indication and a chosen change in count rate set to give full scale deflection.

4. Choice of the Radio-active Source

Bremsstrahlung may be translated from the German as "slowing down or braking radiation" and is the name given to the radiation generated when beta particles are slowed down in the Coulomb fields of atomic nuclei, causing the emission of an electromagnetic radiation with characteristics similar to low energy x and gamma radiation. This radiation exhibits a continuous energy spectrum, the maximum energy being determined by the energy of the exciting beta radiation and the mean energy and intensity by the atomic number of the 'target'.⁴

Consideration of the mass absorption co-efficients shown in Fig. 2 indicates that a radiation energy of less than 30 keV will give the increased sensitivity to sulphur content changes desired. Isotopes such as Fe⁵⁹ and Cd¹⁰⁹ emitting x-rays at 6 and 22 keV respectively, have been used in this type of equipment, but are relatively expensive and short lived compared to the 12.3 year half-life tritium-zirconium source with a mean energy of approximately 8 keV. The efficiency of production of bremsstrahlung being comparatively low, a large amount of the exciting beta isotope is required to produce a source of useful intensity; however, tritium is now available at comparatively low cost and techniques are known which allow the preparation of tritium bremsstrahlung

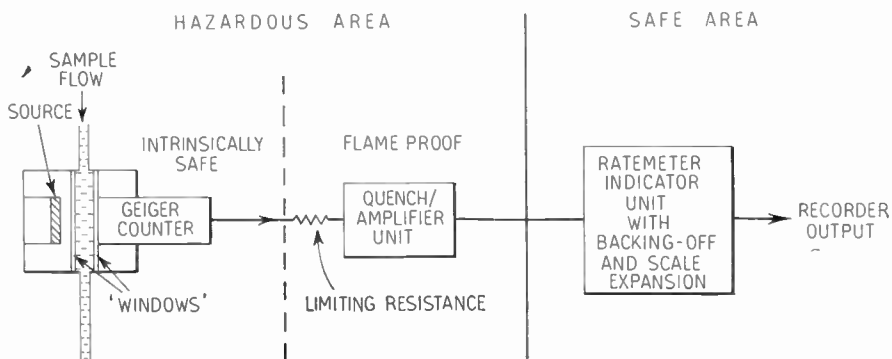


Fig. 3. Schematic diagram of sulphur analyser.

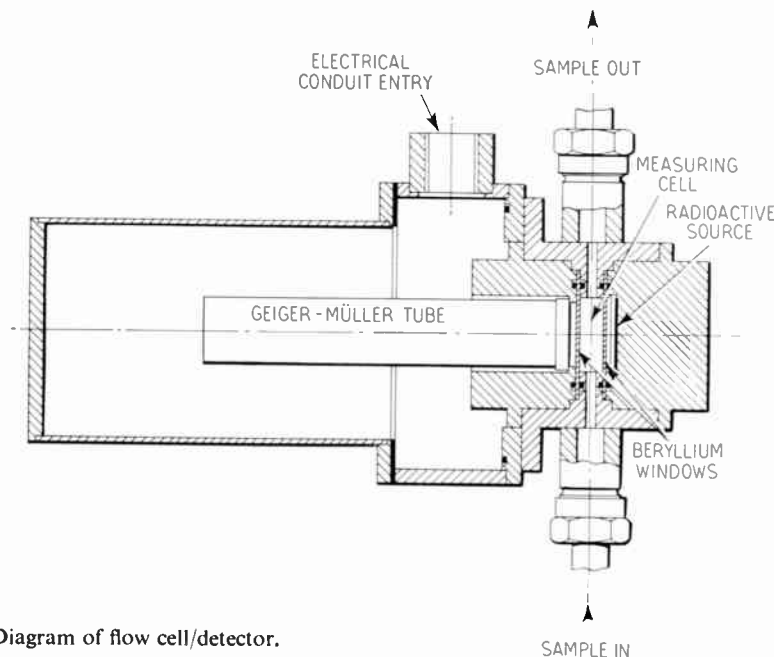


Fig. 4. Diagram of flow cell/detector.

sources containing several curies of tritium. The source is prepared by melting a thin layer of zirconium on to a tungsten disc approximately 1 inch in diameter and heating this 'target' in an atmosphere of tritium; absorption occurs as the disc is allowed to cool. A typical target will take up 1 cm³ of tritium which may have a specific activity of 2.6 curies/cm³. A second bremsstrahlung source produced from promethium 147 in aluminium is available, giving a mean radiation energy of approximately 26 keV and having a half-life of 2.6 years.

These two sources are the only bremsstrahlung sources with mean radiation energy of less than 30 keV and useful half-life at present available. The tritium source has the advantages of long life and low cost compared to the promethium source, while the latter has the advantage of a radiation energy which allows the design of a measuring system where sulphur content indication is, to a large degree, independent of density and carbon/hydrogen ratio of the sample.

5. Construction of the Flow Cell

It will be appreciated that the use of very low radiation energy precludes mounting the source and detector across a standard pipe as in the case of the density gauge; the absorbing mass of the pipe walls would reduce the detected radiation to a prohibitively low level. Refinery practice also required that the unit be mounted on a by-pass from the main stream, so that maintenance and calibration would not interfere with plant operation. A flow cell was designed (Fig. 4) in which the radiation 'windows' were com-

posed of beryllium discs 0.020 in thick. The very low mass absorption co-efficient of beryllium gives minimum loss of radiation through the cell, but the 'windows' are sufficiently strong to allow stream pressures up to 200 lb/in² to be handled. The sample chamber width of approximately $\frac{3}{8}$ in was chosen as a compromise between the maximum useful count rate at the detector from the source intensity available and the desire to inspect as large as possible a sample from the main stream. Mounting arrangements to carry the source and detector adjacent to the cell 'windows' were incorporated into the unit, with the necessary small amount of radiation shielding, to reduce the radiation level on the outer surface of the assembly to a level acceptable under the Ionising Radiations Regulations.

6. Choice of Radiation Detector

Three radiation detectors are commonly available for use in industrial applications: the Geiger-Müller tube, the ionization chamber and the scintillation counter. No ionization chamber of suitable characteristics was available for efficient use at the radiation energy involved in this application, so that the choice of detector was limited to a G-M tube sensitive to low energy radiation or a scintillation counter. The main factors governing the choice of detector for this application are summarized in Table 1 with appropriate comments for each type.

It will be seen from these factors that the G-M tube MX118 suffers from only two disadvantages: a comparatively low detection efficiency and the consequent

long time required for an accurate assessment of count rate. In the refinery applications under consideration, sulphur content changes were expected to occur very slowly so that these points were not considered unacceptable; it was found that with an integrating time of 6 minutes in the ratemeter circuit, the detection accuracy was adequate to enable a change in sulphur content of 0.01 % to be detected.

Table 1

Characteristic	G-M Tube†		Scintillation Counter
	MX 118	MX 122	
1. Relative detection efficiency	1	8	5
2. Suitability for use in hazardous areas	Can be used with an intrinsically safe circuit.		Cannot easily be used with an intrinsically safe circuit.
3. Temperature stability	Good	Poor	Very Poor
4. Electrical stability	Good	Good	Poor
5. Mechanical robustness	Fair	Fair	Fair
6. Associated circuits	Simple		Complex
7. Automatic standardization required	No		Yes
8. Time required for accurate assessment of count rate	Long	Short	Short
9. Cost	Low		High

† Mullard Ltd.; MX118 argon filled, MX122 krypton filled.

7. Measuring Circuits Associated with the Detector

The quench amplifier unit is housed in a 'certified flameproof enclosure' within a few feet of the G-M tube and houses a limiting resistor in the h.t. supply line to the tube. The value of this resistor is chosen so that the maximum charge under fault conditions is insufficient to initiate a spark in the gases involved; thus the detector and its connecting cable are an intrinsically safe assembly. Although the MX118 is a

self-quenching tube and would operate without an external quench unit, the use of this circuit improves both the length and slope of the plateau. The most important function of the quench unit, however, is to lengthen considerably the working life of the tube. This effect is achieved by producing a quenching pulse having a very short rise-time, immediately a discharge is initiated in the tube by radiation, so that the amount of charge taken from the tube per pulse is very much reduced. As it is required that the detector pulses are transmitted through several hundred feet of cable to the ratemeter, an amplifier is incorporated with the quench unit to raise the output pulses to a suitable power level.

Figure 5 shows schematically the ratemeter circuit used to shape and integrate the pulses and to provide an analogue output. The pulse former produces a pulse of fixed amplitude for every incoming pulse; these pulses are fed to the ratemeter where they produce a charge on C1; the resistor connected in parallel with C1 serves to discharge the capacitor at such a rate that the voltage appearing across the capacitor is the analogue of the mean pulse rate during the chosen averaging time constant. A separate circuit with an associated variable control allows the averaging time constant to be altered to suit plant operating conditions. The voltage appearing across the capacitor is measured by the high input impedance valve-voltmeter which provides an output signal suitable for feeding either to a simple pen recorder or to a self-balancing potentiometer recorder. It will be appreciated that change in count rate is of greater interest than actual count rate and the backing-off control allows the ratemeter to be adjusted so that any given input produces zero output. Thus the count rate due to any given sulphur content may be set to give a zero reading on the recorder. The scale expansion control adjusts the deviation sensitivity to allow a given change in count rate to be set to give full scale deflection on the recorder. Finally, the sensitivity control adjusts the sensitivity of the pulse former so that while the number of counts received for any given sulphur content will decrease with time due to the

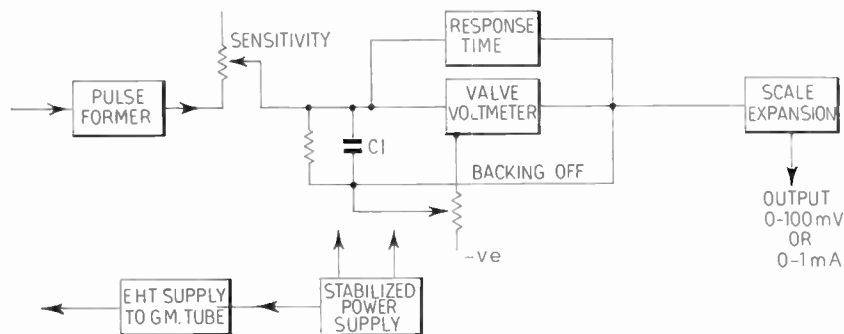
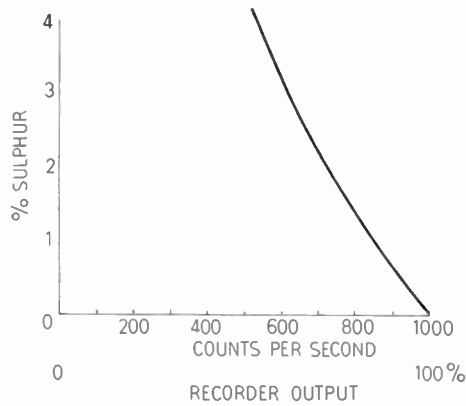
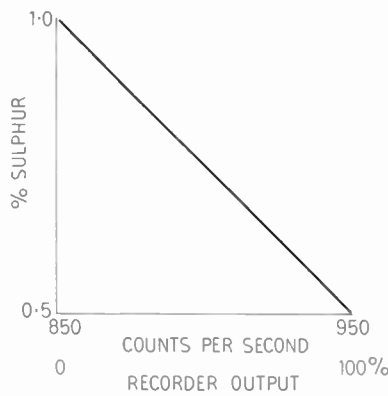


Fig. 5. Schematic diagram of ratemeter circuit.



(a) Total detected counts.



(b) Backed-off counts.

Fig. 6. Relationship between count rate and sulphur content.

natural decay of the source, the amount of charge passed to the ratemeter per count can be increased to compensate for this effect and maintain an accurate calibration.

8. Accuracy of Measurement

The limit of accuracy to which sulphur content can be indicated in a given sample rests mainly upon three factors; the sensitivity of the detection system to small changes in radiation level; the statistical accuracy to which the fluctuating detector signal can be assessed and the stability of the electronic circuits. The sensitivity of the detection system is defined by the radiation energy and the geometric layout of the flow cell, but the indication sensitivity can be increased by using the backing-off and scale expansion facilities of the ratemeter. Figure 6(a) shows the relationship between detected counts per second and sulphur content of a given sample using a standard ratemeter; the ratemeter signal below about 500 counts per second is clearly of no interest in this application and only a small portion of the ratemeter output covers a

large sulphur content range. Figure 6(b) shows the type of relationship between sulphur content and indicator reading which may be obtained by backing-off the unwanted part of the signal and spreading the count rate change due to a particular sulphur content change over the full scale of the indicator.

The random nature of the emission of radiation from the source imparts a statistical fluctuation to the ratemeter output so that it is necessary to use fairly long averaging times in order to obtain an accurate assessment of the count rate. While the ratemeter is accurate to $\pm 0.1\%$ of count rate under ideal conditions, the accuracy expressed in terms of sulphur content of $\pm 0.005\%$ which this ratemeter accuracy suggests from Fig. 6(b) is not achieved in practice. The accumulation of small errors in reading due to electrical pick-up, heavy mains surges, counter stability etc. limits the overall accuracy to about $\pm 0.02\%$

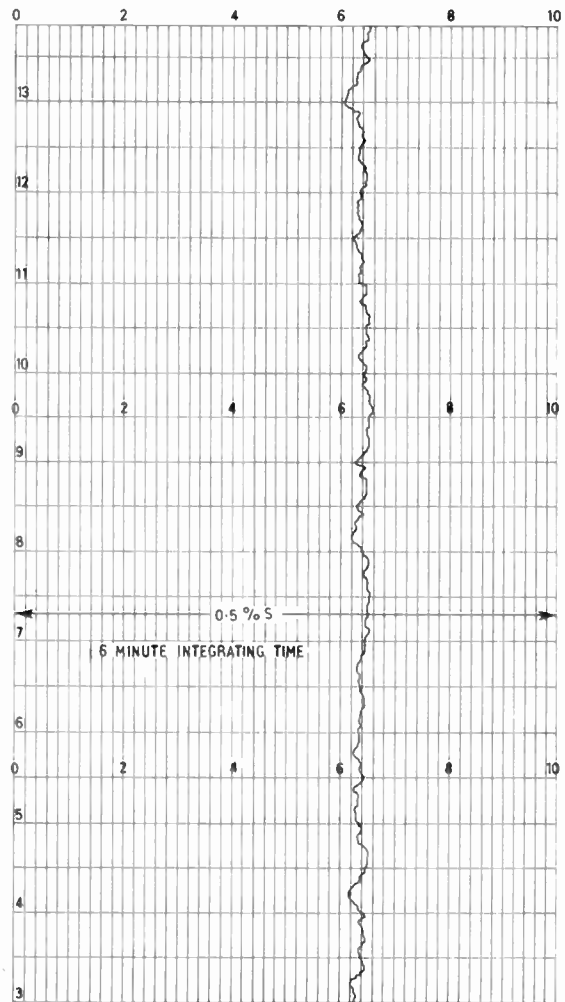


Fig. 7. Recorder trace showing statistical variation of signal compared to magnitude of signal change for actual sulphur content change.

sulphur, in applications where due correction can be made for the effects of variation in density and carbon/hydrogen ratio. Figure 7 shows a typical recorder trace from an equipment set up to cover a range of 0.5% sulphur and it will be seen that although the mean sulphur level can be estimated to about $\pm 0.01\%$ the overall width of the trace is approaching $\pm 0.02\%$ sulphur.

Possible causes of error in reading sulphur content with this type of equipment become apparent on further study of the mass absorption co-efficients shown in Fig. 2. Firstly, the presence of elements other than hydrogen, carbon and sulphur in appreciable or varying quantities in the stream will obviously give rise to errors; especially if they are elements of high atomic number. Secondly, although the major radiation absorption change will occur as a result of sulphur content change, a significant change in density or carbon/hydrogen ratio will cause a change in the total sample absorption characteristic. It has been found that the effect of density change expressed as an error in sulphur reading, varies between different types of base stock and can be as much as 0.09% of sulphur for a 0.001 s.g. change. It is therefore essential that density, carbon/hydrogen ratio and contaminating element concentration are known or sensibly constant if this method of sulphur determination is to be used. It is also found that changes in the temperature of the flow cell give rise to changes in indicated sulphur content and in order to obtain the overall accuracy figure of $\pm 0.02\%$ it is necessary to maintain the cell temperature constant to within ± 10 deg F.

9. Sulphur Determination in Laboratory Samples

The sensitivity to sulphur content changes of a system of the type under discussion may be expressed as:

$$S = W_s(u_s - u_{ch})m$$

where S = fractional change in transmitted radiation for a given fractional change in sulphur content

W_s = the proportion by weight of sulphur

u_s = the mass absorption co-efficient of sulphur

u_{ch} = the average mass absorption co-efficient of the sample

m = the mass/unit area of the sample.

It will be seen that for S to be dependent only on changes in W_s , both u_{ch} and m must be constant, or vary insignificantly in the particular application as is the case in certain refinery streams. In a constant thickness cell, m will obviously vary in proportion to the density of the sample and to maintain m constant, some form of constant weight cell must be used. u_{ch} depends on the radiation energy and the carbon/hydrogen ratio of the sample if contaminating elements are present in only insignificant quantities.

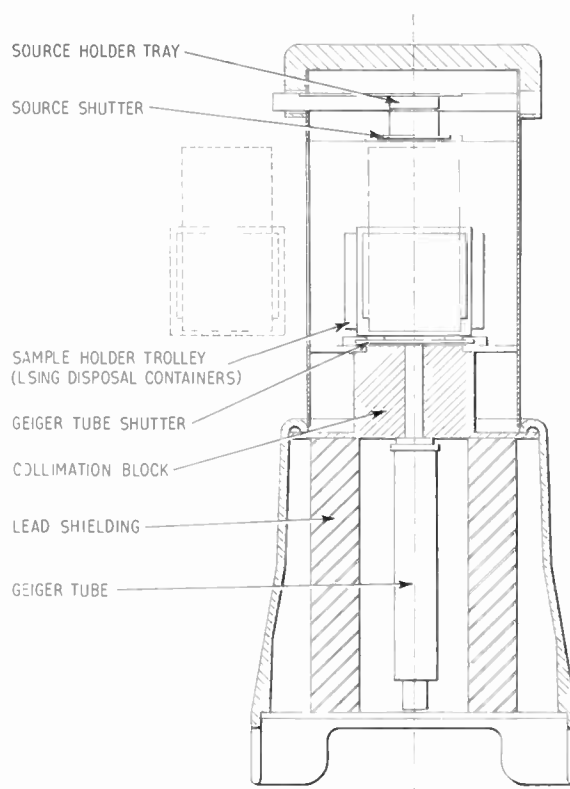


Fig. 8. Diagram of laboratory sample analyser.

At a radiation energy of 21.4 keV the values of the mass absorption co-efficients for hydrogen and carbon are found to be equal under certain narrow beam geometrical conditions of source, sample and detector. Although a bremsstrahlung source of 21.4 keV is not known, the $\text{Pm}^{147}/\text{Al}$ source having a mean energy of about 26 keV can be used in a system which closely approaches the ideal for sulphur determination independent of density or carbon/hydrogen ratio variations. The use of a detector whose sensitivity falls off rapidly with increase in radiation energy effectively reduces the mean detected energy. The use of a sample mass of appropriate value to filter out the low energy components of the radiation effectively raises the mean detected energy. Finally, the degree of collimation of the radiation beam affects the sample mass absorption co-efficient and the energy at which hydrogen and carbon have equal mass absorption co-efficients.

Using these three factors it is possible to produce an arrangement as shown in Fig. 8, where constant weight samples in standard polythene containers may be examined in four to five minute periods to give sulphur content readings accurate to $\pm 0.04\%$ over the range 0.5% to 2.5% sulphur independent of changes in carbon/hydrogen ratio.

The indicator used in conjunction with the detector assembly shown in Fig. 8 is an automatic scaler which displays the number of pulses received in a preset time or the time taken to receive a preset number of pulses. As the statistical accuracy of the count is dependent upon the number of events recorded, the latter system is favoured as constant statistical accuracy is achieved.

Although the readings obtained may be independent of carbon/hydrogen ratio over a certain range of sulphur content, the change in sulphur content of the sample will itself cause a small change in the average mass absorption co-efficient of the whole sample and thus it is necessary to adjust one of the three parameters governing carbon/hydrogen ratio independence. The sample mass can easily be changed to give the best independence over the range of sulphur concerned. Taking into account errors due to G-M tube and electronic instability and difference in carbon/hydrogen ratio such as are found between benzene and hexane, overall errors in sulphur content indication have been found to be $\pm 0.03\%$ in the range 0.5% to 1% sulphur with 60-gramme samples and $\pm 0.05\%$ in the range 2.5% to 6% sulphur with 35-gramme samples.

While the equipment described in this paper was designed specifically as a result of the requirement to determine the sulphur content of hydrocarbons in a refinery, consideration is being given to other applications such as the measurement of tetra-ethyl lead content of petroleum fuels and sulphur dioxide in an organic base. It is likely that as further applications

are examined, the scope of this simple analytical tool will be widened appreciably.

10. Acknowledgments

The author wishes to record his gratitude to Mr. H. W. Finch and Mr. A. W. Abbott for their assistance in the preparation of this paper and to the Directors of Isotope Developments Limited for their permission to publish it.

11. References

1. R. Y. Parry and A. E. M. Hodgson, "Fluid density measurement by means of gamma ray absorption", *J. Brit.I.R.E.*, **17**, p. 365, July 1957.
2. G. Syke, "Radio-isotopes in metal gauging", *Nuclear Engineering*, **2**, February 1957, p. 72.
3. G. Syke, "Gamma-ray thickness gauge for hot steel strips and tubes", *J. Brit.I.R.E.*, **14**, p. 419, September 1954.
4. J. F. Cameron, J. R. Rhodes and P. F. Berry, "Tritium Bremsstrahlung and its Applications", U.K.A.E.A. Research Group Report. A.E.R.E. No. R3086, 1959.
5. R. E. Pegg and J. S. Pollock, "Continuous Analysis of Sulphur in a Refinery", Proceedings of the Conference on the Uses of Radioisotopes in the Physical Sciences and Industry, Paper RICC/11. Copenhagen 1960.
6. J. R. Rhodes, T. Florkowski and J. F. Cameron, "Analysis of Sulphur and Cobalt in Hydrocarbons using a $^{147}\text{Pm}/\text{Al}$ Bremsstrahlung Source", U.K.A.E.A. Research Group Report, A.E.R.E. No. R3925, 1962.

Manuscript first received by the Institution on 16th January 1963. (Paper No. 853.)

© The British Institution of Radio Engineers, 1963

DISCUSSION

Mr. W. H. Topham: An x-ray absorption sulphur monitor, developed by BP in co-operation with Saunders Roe and Nuclear Enterprises Limited, has been used successfully at Kent Refinery to control the quality of a gas oil stream. The monitor, shown diagrammatically in Fig. A, uses two radio-active sources ($^{147}\text{promethium/zirconium}$), a 3 in bore plastic flow-through sample cell, and a scintillation counter detector with 5 in diameter phosphor. To overcome source decay and detector drift, the measuring system is standardized every 12 minutes by operating the reversing motor to obscure the measuring source and expose the detector to the reference source. The standardizing technique is similar to that used on battery-operated potentiometric recorders, but utilizing a radiac source in place of a standard cell. The instrument housing is air-pressurized for safe operation, with pressure switches. Accuracy and reliability are comparable with the equipment described by Mr. Rowley. On this particular application, the density of the sample was monitored separately, and sulphur contents calculated from the two measurements. The effect of C/H ratio is very small with a promethium source.

New monitors, of much higher sensitivity and indepen-

dent of C/H ratio and density, are currently being developed for sulphur and lead applications in refineries and blending installations.

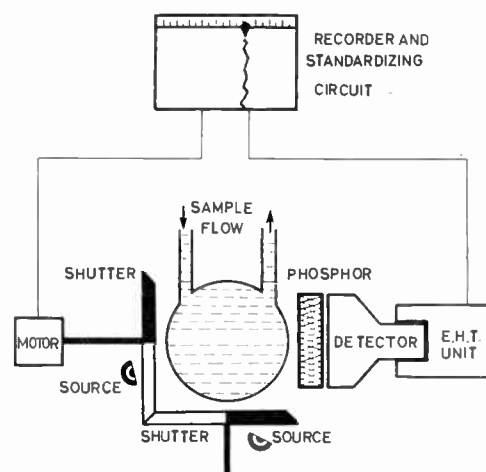


Fig. A. Diagram of prototype sulphur monitor.

Measurement of Phase Shift at Audio Frequencies using a Magslip as a Calibrated Reference Standard

By

A. G. J. HOLT, Ph.D., Dip.El.

(Associate Member)†

Summary: A method of phase measurement is described which does not appear to have been used before at frequencies in the audio range.

The general principle is stated, the apparatus is described and a curve comparing the measured and predicted phase response of a simple network is given.

1. Introduction

Many situations require the determination of the relative phase displacement between two sinusoidal voltages but commercial equipment is frequently unsuitable or not available. In the author's experience such a situation has occurred when carrying out measurements on active and passive filter networks.

Of the many methods of phase measurement which have been described^{1, 2, 3, 4}, those which depend upon the geometrical properties of patterns traced upon the screen of a cathode ray tube are of doubtful practical value, inconvenient in operation and of low accuracy, whereas others employ equipment that is expensive and complicated.

The method to be described is simple and avoids the use of carefully calibrated components to give a reliable indication of phase shift. This is accomplished by employing a magslip as the phase shifting component which can be used for phase measurements over a range from zero to 360 deg or more without switching at say ± 90 deg or ± 180 deg, as is required in some forms of equipment.

A c.r.o. is used but only as a visual null indicator and at no stage is it necessary to perform calculations on, or measurements of, the observed pattern.

2. General Principle

Figure 1 shows a block diagram of the calibrated phase shift equipment.

The phase shifter employs the principle that a rotating magnetic field is produced by currents flowing in coils whose displacement in space is equal to their displacement in phase. The two identical stator windings of the magslip are wound with their axes at right angles and are supplied with voltages differing in phase by 90 deg. The rotating magnetic field which results, induces in the rotor winding an e.m.f. whose magnitude is independent of rotor position but whose phase is directly related to the angular setting of the rotor. The magslip‡ is converted to a calibrated phase shifter by attaching a circular scale, calibrated in

degrees, from which the phase displacement can be read directly. This application of the magslip is well known at power frequencies but does not appear to have been applied at frequencies in the audio range. It is clear that successive measurements which may produce a total phase shift of 360 deg or more and might ordinarily be difficult to handle, simply correspond to something more than one revolution of the rotor shaft.

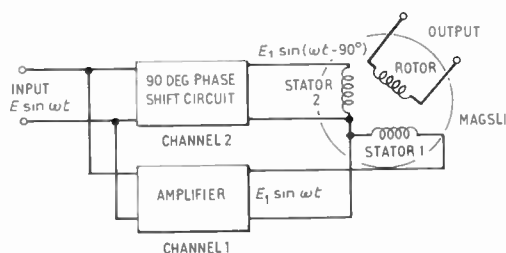


Fig. 1. Block schematic of the calibrated phase shift equipment.

In Fig. 1 stator 1 is supplied from the low output impedance of the feedback amplifier and cathode follower which constitute channel 1. Since the amplifier has an open loop gain of approximately 100, the output impedance of channel 1 is not greater than 10 ohms. Stator 2 is supplied from the 90 deg phase shifting integrator and cathode follower that make up channel 2, which also has a low output impedance. Both output impedances are much less than the impedance of the stator windings which have a measured resistance of 1040 ohms in series with an inductance of 175 mH. It follows that the channel outputs are effectively ideal voltage sources and that the stator winding currents will be in quadrature when the phase voltages differ in phase by 90 deg. It is of course necessary for the currents flowing in the two stator windings to be of equal magnitudes at frequencies within the working range. The gain of channel 1 is almost constant over the frequency range of interest, whereas the output of the valve integrator in channel 2 is not. It is, in fact, inversely proportional

† Department of Electrical Engineering, University of Newcastle upon Tyne.

‡ The magslip used is 3 in. diameter and manufactured to Service Specification AP 10861.

to the frequency of operation as is shown by the appearance of $\omega = 2\pi \times \text{frequency}$ in the denominator when the function $\sin \omega t$ is integrated with respect to time. It follows that the integrator amplifier must be operated on that part of its frequency characteristic where the gain is inversely proportional to frequency, that is, in the range $\omega > \omega_1$ in Fig. 2.

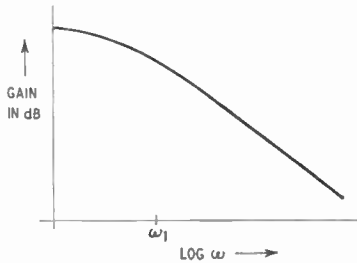


Fig. 2. Frequency vs gain of the integrator amplifier.

Hence if the frequency of operation is changed the output voltage from the integrator is altered and the variable components in the integrator circuits must be adjusted in order to maintain equality of the two-channel outputs.

3. Operating Procedure

When it is used to carry out a measurement of the phase displacement introduced by a network, the apparatus is set up as in Fig. 3.

In the absence of harmonics, an ellipse will ordinarily appear on the screen of the c.r.o., except for signals which are cophasal, when a straight line will be seen. Thus by setting switch S to position (1) and adjusting the mag slip rotor until a linear c.r.o. trace is observed, the zero reference point on the rotor scale is found. Subsequently, with S in position (2) a further rotation is required to reproduce a linear trace and the angular displacement is equal to the phase shift introduced by the included network. Greater accuracy is achieved⁴ by increasing the gain of the c.r.o. amplifiers so that only a small part of the ellipse can be accommodated on the screen. A precise adjustment is then obtained by turning the rotor until the two observed parallel lines are coincident. Note that at no time is it necessary to measure the dimensions or angles of any patterns appearing on the trace. This eliminates errors due to the curvature of the tube face and parallax errors which occur when a graticule is used.

4. The Calibrated Phase Shifter Circuit

A circuit diagram of the calibrated phase shifter is shown in Fig. 4.

Here, V1(a) is a cathode follower which isolates the following amplifier and integrator from the effects of

the output impedance of the signal generator. A decoupling circuit is connected in the anode circuit of the triode valve to prevent interference over the high tension supply from entering the input. Channels 1 and 2 are fed from the cathode follower output.

Channel 1 consists of pentode amplifier V2 and a cathode follower using both halves of a double triode V3 double triode to provide ample current for the stator coil at all frequencies of operation. Resistors RV1 and RV2 are used to apply heavy overall negative feedback from the cathode follower output to the control grid of the pentode. These two resistors are used to control the gain of the amplifier stage. A large blocking capacitor is used between the cathode follower and the mag slip stator winding.

Channel 2 also consists of a pentode amplifier V4 and both halves of V5 used as a cathode follower. A capacitor C1 is used to provide feedback between the cathode follower output and the control grid of the pentode. Two carbon track variable resistors RV3 and RV4 are connected between the output from the isolating cathode follower V1(a) and the control grid of the pentode. These resistors, in conjunction with the feedback via capacitor C1, ensure that for sinusoidal waveforms the output voltage from the cathode follower V5 is the time integral of the voltage at the input terminals. Because the feedback in channel 2 will vary with frequency and because the magnitudes of the stator currents must be the same, the output of channel 2 must be adjusted as the frequency changes. Resistors RV3 and RV4 allow the magnitude of the output voltage to be set at a predetermined level, equal to the output voltage from channel 1, at each frequency of operation.

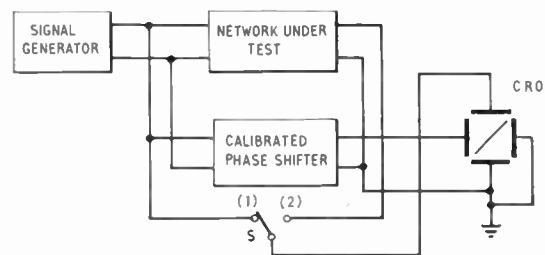


Fig. 3. Block schematic of the equipment used for phase displacement.

In order to facilitate the matching of the channel output voltage, a simple valve voltmeter is provided in the equipment. This consists of half of a double triode V1(b) arranged to have both anode and cathode loads and to be connected to four transistor diodes arranged as a bridge. A moving coil milliammeter of 1 mA full scale deflection is connected between the midpoints of the bridge. The undecoupled cathode

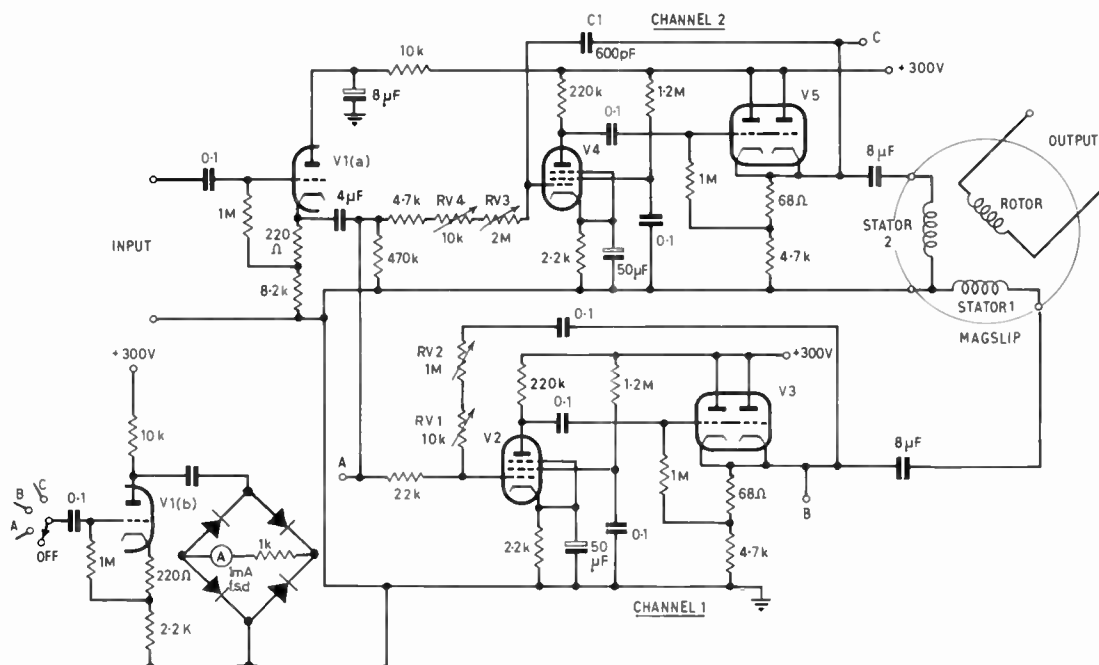


Fig. 4. Circuit diagram of the calibrated phase shifter.

resistor allows a larger amplitude voltage to be applied to the voltmeter than could be applied if a decoupling capacitor were fitted. A four-position switch S1 is provided to allow the valve voltmeter to be connected to point A at the output from the isolating cathode follower V1(a) or to either of the magslip stators at points B and C. It is convenient to set the voltages to all these points to the same value. This eliminates the need for calibration of the valve voltmeter since the input voltage from the signal generator can first be adjusted to give a suitable deflection on the milliammeter with the selector switch on position A. The gains of channels 1 and 2 can be set to give the same meter deflection when the selector switch is in positions B and C by adjustment of the variable resistors in the channel circuits. Thus setting the gains to unity ensures low output impedance from both channels.

The valve voltmeter can be dispensed with if the outputs at switch S1 are taken to the c.r.o. It is then only necessary to adjust the signal generator input and the channel gains to give the same pattern amplitude between two marks on the c.r.o. graticule.

A 300 V stabilized power unit is used to supply high tension and heater power for the equipment.

5. Setting Up Procedure

To set up the circuit a voltage at the required frequency is applied to the input terminals of the isolating cathode follower and the valve voltmeter in

the equipment is switched to position A (Fig. 4). This allows the input voltage to channels 1 and 2 to be indicated. The amplitude of the applied voltage is adjusted until the milliammeter in the voltmeter circuit indicates one half of full scale deflection (0.5 mA). This corresponds to 1.9 V r.m.s. at the point A. The voltmeter is then switched to point B, which is connected to the magslip stator winding supplied from the amplifier channel. The settings of the feedback resistors RV1 and RV2 are adjusted to give one half of full scale deflection on the meter, when the gain of channel 1 is unity. The voltmeter is next switched to the output from the integrator channel at the point C which is connected to the other magslip stator winding. In order to set the gain of this channel to be equal to unity the resistors RV3 and RV4 must be adjusted until the milliammeter again indicates one half of full scale deflection. When these adjustments have been made the circuit is calibrated and ready for use.

If the frequency of operation is changed it is advisable to check the voltage levels again. In practice it is found that the input voltage and the gain of the amplifier channel require very little adjustment over a wide range of frequencies. It is usually necessary to carry out adjustments to the integrator channel at each different frequency. When the frequency of operation is changed the settings of the resistors R3 and R4 must be adjusted if the gain of the integrator channel is to be maintained at unity.

6. Performance

A test of the accuracy of the apparatus was carried out by comparing the measured phase/frequency response of a network with that obtained graphically from a plot of the pole locations of the network function. The network used was a low pass filter designed to the three-pole approximation to the maximally flat delay characteristic. This filter was designed to approximate to a linear phase/frequency response in the pass band from d.c. to 4 kc/s. A design table for filters of this type has been published by Weinberg.⁵ Writing the network function

$$N(p) = \frac{\text{output voltage}}{\text{input voltage}}$$

$$\text{gives } N(p) = \frac{H}{(p + 1.325)(p + 1.05 + j1)(p + 1.05 - j1)}$$

where H is a constant and p is the Laplace transform variable, $p = \sigma + j\omega$.

A plot of the pole locations is given in Fig. 5.

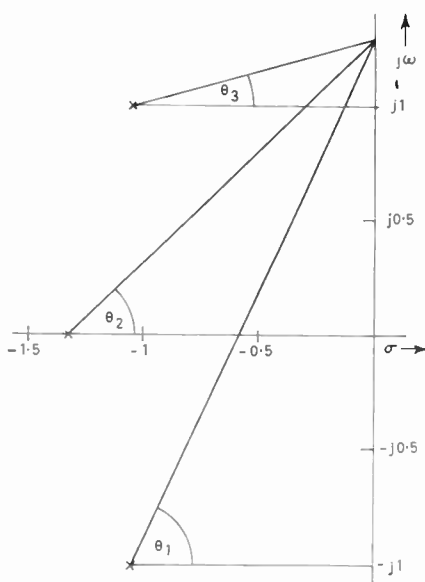


Fig. 5. Plot of the pole locations.

At any angular frequency ω , the phase shift $\theta(\omega)$ is given by $(\theta_1 + \theta_2 + \theta_3)$.

Figure 6 shows the phase shift obtained graphically together with the results of a series of measurements using the apparatus described in this article. It is evident from the curve that the equipment is accurate to ± 2 deg over the frequency range from 300 c/s

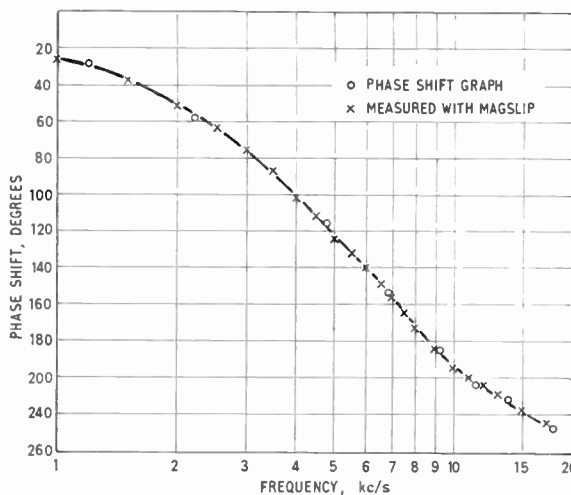


Fig. 6. Graph of phase vs frequency of a given network.

to 18 kc/s. This range of frequencies was greater than was required for the immediate purpose for which the phase measuring equipment was designed so its normal operating range was restricted to 7 kc/s. This range can be extended if the capacitor C_1 in Fig. 4 is reduced to 200 pF.

7. Acknowledgments

The author would like to thank Professor R. L. Russell, Professor of Electrical Engineering in King's College, Newcastle upon Tyne, for the use of the facilities of the Department and for helpful criticisms of the draft of the paper. Thanks are also due to Mr. F. J. U. Ritson, Reader in the Department of Electrical Engineering, for stimulating discussions.

8. References

1. F. E. Terman and J. M. Pettit, "Electronic Measurements" (McGraw-Hill, New York, 1952).
2. F. E. Terman, "Radio Engineers' Handbook", (McGraw-Hill, New York, 1943).
3. J. Fritz, "Precision phasemeter for audio frequencies", *Electronics*, 23, p. 102, October 1950.
4. J. R. Ragazzini and L. A. Zadeh, "A wide-band audio phasemeter", *Rev. Sci. Instrum.*, 21, p. 145, March 1950.
5. L. Weinberg, "Additional tables for the design of optimum ladder networks", *J. Franklin Inst.*, 264, p. 127, August 1957.

Manuscript received by the Institute on 9th January 1963. (Contribution No. 70.)

© The British Institution of Radio Engineers, 1963

Man and Machine in the Extraction and Use of Radar Information

By

R. BENJAMIN, B.Sc., A.C.G.I.†

Presented at a Symposium on "Processing and Display of Radar Data" arranged by the Radar and Navigational Aids Group and held in London on 16th May 1963.

Summary: Modern primary and secondary radars can potentially provide far more information than can be directly displayed—or can be absorbed by the user. Hence there is a strong *prima facie* case for automatic data extraction. Even the filtered information arising from automatic extraction cannot be efficiently used by the unaided human operator.

However, the use of a surveillance radar picture in air or sea traffic control or in military tactical control involves a vital element of human appreciation, judgement and policy formulation. Fortunately, the remaining aspects of these tasks can mainly be defined as routine applications of pre-determined doctrine and are largely suited to automatic operation. Hence automation of both data extraction and the routine aspects of data utilization may be required, to free man to make full use of his special talents.

If the radar is not mechanically constrained to distribute its power and attention quasi-uniformly over the solid angle covered, considerable gains could be derived from automatically matching the operation of the radar to the data emerging from the equipment.

A research and experimental programme has been conducted over some years, aimed at the application of automation to radar data processing and, eventually, the automatic control of the radar's mode of operation. The paper outlines some of the requirements and problems in matching the man and machine to each other, to their input sources and to the appropriate output systems.

1. Introduction

New sensors and communication channels are ever increasing the already overwhelming quantity of information available to the user of radar information. On the other hand, the time available for decision is ever decreasing. We must therefore exploit all possible advances in visual and auditory 'display' techniques to maximize the appreciation capability of the human operator. However, even the best matching of the input to the human being cannot bridge the gulf of several orders of magnitude between his data-processing 'bandwidth' and that of the incoming information. Hence the foremost requirement is clearly to devise automatic means of extracting from our data sources the relevant signals. The resultant information must then be automatically processed, in accordance with appropriate operational doctrines, so as to distil from it that irreducible essence which is:

- (a) Required for the correct evaluation and use of the total ensemble of information,

- (b) Not amenable to automatic processing,
 (c) Well matched to human operator's unique powers of pattern-recognition,
 (d) Well within the operator's processing capacity.

The choice of *what* to display, and the selective gathering of that information, is thus even more important (and probably more difficult) than the decision *how* to present this information and the design of the appropriate 'hardware'.

The creation of an efficient partnership between man and machine presents some challenging problems in the development of highly skilled, industrious and adaptable robots. However, it also presents no less difficult psychological and organizational problems and can probably only be solved by the engineer, scientist and prospective operational user working as a single, dedicated but self-critical team.

Researches in this field, over the last few years, have underlined the extent to which the tacit acceptance of human limitations in data extraction and processing, and in display utilization has set limits on the design of sensing sources, data links and control

† Admiralty Surface Weapons Establishment, Portsmouth, Hampshire.

organizations, and has prevented the human operator from playing his full and proper part in the operational use of radar information.

2. The Relative Capabilities of the Radar Receiver, Display and Operator

2.1. *Intrinsic Radar Capacity*

Consider a radar with a 1- μ s range discrimination and a 10-s aerial rotation period. Assume that this radar can receive echoes at all ranges, up to the limit set by the pulse-repetition rate. If the radar can just distinguish differences in angular position corresponding to one azimuth pulse spacing, it will survey 10^7 discriminable potential echo positions, in plan. This result is independent of the pulse repetition rate, and arises from the fact that in the rotation time of 10 seconds the equipment receives 10^7 independent amplitude samples.

Now assume that the radar has eleven beams, of characteristics equivalent to the above, stacked vertically one above the other. In each of the ten regions, between two adjacent beam centres, roughly ten fine-elevation positions should then be distinguishable, on the basis of the relative amplitude of the echoes in the two receiving beams concerned. Each possible plan position is thus associated with 100 distinguishable elevation positions. Finally assume that a form of 'swept gain' is used to compensate for the inverse square law of radiation from the transmitter and re-radiation from a target, by multiplying the signal (and noise) emerging from the receiver by the fourth power of the range. The resultant signal output is then solely a function of a target's 'echoing area'. This echoing area, averaged over all the pulses in the beam-width, can then be classified into, say, ten distinguishable amplitude brackets. Thus each revolution of the aerial system would give information on the presence of targets in 10^{10} discriminable positions in the four co-ordinates of range, bearing, elevation and target echoing area.

Each actual target overlaps roughly ten discriminable increments in each of the dimensions of bearing, elevation and amplitude. Thus the 10^{10} intrinsically discriminable elements of (generalized) 'position' can accommodate only about 10^7 intrinsically resolvable targets, each described by ten binary digits defining its 'fine' position within the thousand discriminable elements making up one 'resolution box'. Hence the total system information capacity is 10^8 bits per radar rotation.

2.2. *Comparison of Capacities*

A p.p.i. display might have a 'definition' of 800 television lines, giving 5×10^5 discriminable spot positions, and it might permit potential targets to be classified as 'large', 'medium', 'small', or 'invisible'.

Thus the 10^{10} elements in four dimensions are compressed, for display, into 2×10^6 elements in two primary dimensions. (In this process significant echoes may become submerged in signals, clutter or noise corresponding to positions in space which are resolvable from the wanted target in the radar output, but not on the display.) If the average echo occupies ten spot positions and its centre can be judged to an accuracy of roughly one spot diameter, we have 5×10^4 resolvable potential echo positions, each associated with three binary digits describing its fine azimuth position and two describing its amplitude. Hence the total display information capacity (neglecting afterglow) is roughly 2.5×10^5 bits per radar rotation.

The practical human operator might be credited with a capacity of about four independent binary recognition decisions per second (i.e. 40 per radar rotation). If this were all the story, one man would require two eight-hour days to take in one 'frame' of a p.p.i., and he would take three hard working years to take in the full information from one rotation of the radar.

Fortunately, we know that only very few of the 10^7 potential echo positions will in fact be occupied by meaningful targets, and these targets will be subject to limitations on their possible positions, signal amplitudes, velocities and accelerations. Moreover, the targets of interest will largely arise in groups conforming to a recognizable, consistent and meaningful pattern of behaviour. The combination of man and display can exploit some of this redundancy to minimize the disparity in surveillance capability indicated above, but clearly there is little hope of overcoming a handicap represented by the difference between three strenuous working years and ten seconds. Hence, if we wish to exploit anything like the potential target-handling capacity, data rate, accuracy and resolution of modern radar systems, we must develop appropriate automatic data-extraction systems.

2.3. *Comparison of Detection Performance*

In principle, a well-designed machine should be able to do somewhat better than the combination of man and display, in picking out all patches of received signal exceeding their local noise environment by a given threshold, and in rejecting those signals whose amplitude—or amplitude distribution in range and bearing—is not consistent with the type of target looked for. (This should be so, ideally, even in the absence of the resolution compressing effect of the p.p.i. discussed in the preceding section.) However, the machine must tentatively accept *all* such potential targets, in order to build up the track information to which the criteria of plausible behaviour can then be applied. The man, on the other hand, can defer acceptance until the afterglow track, built up by the

first few echoes, gives him a somewhat crude indication of the consistency and motion of the potential target. Fortunately, a machine can, in general, be made to cope with the maximum likely rate of input of potentially significant signals. However, in the absence of Doppler velocity discrimination, it will be unable to detect, say, a small target in the midst of bird-migration returns of similar characteristics. Following first detection, an auto-extraction system can measure most of the redundant features quite accurately, and use them precisely to compile an accurate plot of the position of the various objects of interest.

Automatic devices are weak in recognizing and interpreting patterns relating groups of associated echoes. Here man can play a valuable part, by adjusting the operation of a machine to take account of particular target patterns, echo characteristics, clutter areas or other special aspects of the signals, or of the operational environment.

2.4. Other Data Sources

2.4.1. Secondary radar

A machine can fairly readily note and memorize the response codes received from each target for each secondary radar interrogation mode in use. If secondary radar responses were used to indicate airframe identity, height, etc., this facility could make a revolutionary contribution to track identification (or re-identification), to the maintenance of track continuity, to accurate tracking in height, and to the detection on the ground of malfunctioning of an aircraft's altimeter. Similarly, the automatic identification of ships' echoes and possibly coding of their courses, speeds and intentions, by secondary radar, could substantially assist marine navigation, by indicating the identity, characteristics and possibly manoeuvres of the vessel represented by a given radar echo, and by permitting a vessel of given plan position to be addressed selectively on a suitable communication circuit.

2.4.2. Passive reception and direction finding

Voice transmissions, digital links and other emissions from the vessels of interest provide a flood of potentially useful information which, in its raw form, would overwhelm the human operator. In principle, automatic devices can

- (a) detect these transmissions,
- (b) note their nature,
- (c) note their direction,
- (d) use them for triangulation, in conjunction with similar radiations received by co-operating stations,

- (e) associate them with tracks built up from primary and/or secondary radar, and
- (f) note any significant changes in the presence or nature of these transmissions.

Thus the operationally significant aspects of this type of information can be filtered out, and displayed or recorded as appropriate.

2.4.3. Data links

The appreciation of the operational situation, as it affects any one station, must be based on knowledge of the information obtained, the action taken and the operational capabilities of neighbouring stations. Hence, any automatic radar data processing system must in general include the transmission and reception of information by a variety of (mainly) digital data links. The computers at each station can then select the right data at the right time for transmission, or for accepting them from a link, in each case performing the appropriate translation, addressing and error detection and correction operations.

2.5. Automatic Control of Radar Operation

If the best use is to be made of the transmitter power and receiver bandwidth of a radar system, the 10^7 resolvable space elements referred to in an earlier paragraph should not all be illuminated with equal power levels or investigated with equal frequencies of 'look'. A quasi-inertia less radar system could permit large economies in the use of the radar's potential, by exploiting some of the redundancies previously referred to. This entails both control of the radiated power, as a function of the angle of 'look' (in two dimensions) and as a function of time, and control of the direction and range limits covered by the receivers and data extraction system, at each instant of time. All these characteristics should be continuously adjusted so as to match:

- (a) the instantaneous interference environment,
- (b) the instantaneous echo amplitude from each target,
- (c) the geometric distribution of target positions and velocities,
- (d) the instantaneous requirements of track continuity and plotting accuracy, and
- (e) the needs of any control system, during the current phase of its operation.

The policy for the deployment of the radar's power and capacity would then be formulated by the user, and inserted into the system in accordance with a suitably flexible pre-planned format. The routine, executive implementation of that policy could then be left to the automatic control system.

3. Automatic Processing of Operational Data by Man and Machine

3.1. Human Limitations in Operational Control

The evaluation of a dynamic situation is mainly a matter of recognizing significant patterns, but in addition to the plan dimension shown on normal displays, these patterns involve height, velocity (in three dimensions) and assorted tabular data on performance limits, fuel states, flight plans, traffic schedules, and so on. Some of these additional items of information can be expressed as suitable codes on plan displays, or can be obtained by cross-reference to and from 'totes'. Others can be assessed by allowing the display to advance in time at a greatly accelerated rate. Such a preview of the future would be based on extrapolation from the vehicles' manoeuvres in the immediate past and/or on proposed or hypothetical manoeuvres of selected vehicles of particular importance. However, this still leaves a very heavy burden on the operator. Under any but the lightest loads, the unaided user has to restrict himself (consciously or otherwise) to the rigid, routine application of a predetermined doctrine, and may cope imperfectly even with this relatively simple, machine-like task.

3.2. The Man-Machine Combination

Any truly efficient control system would recognize the limited 'bandwidth' of the human, and would embody the doctrine or drill for the above routine executive functions in the instructions given to an automatic computer. It would thus give the man the time, as well as the information, to determine his operational policy and to monitor and review its effectiveness. When required, the controller can then revise the basic principles of this policy, the weight given to particular factors in its implementation, or its application to a particular vehicle or local situation.

The machine can assist this process by summarizing or analysing the results achieved. In general, any change in basic policy, during operation, should be unlikely, since this basic policy represents mainly the broad objectives and criteria of desirability. Its application to individual vehicles depends mainly on rapid reliable computation of multi-dimensional relative-velocity problems over a curved earth. These computations may well be further complicated by changes in courses and speeds, and by repercussions, possibly through several interacting stages, on other vehicles. Hence they are normally best left to the machine, though the latter may present the user with a small number of alternative decisions, and show the predicted outcomes and implications in a concise form. However, the most profitable role for the controller is likely to be the use of his extraordinary

adaptability and pattern perception to alter the weights and priorities given to the various considerations and data in the machine's program.

The formulation of the general operating doctrine, its overall evolution, and its temporary or local modification must clearly be the controller's responsibility. Unless there is calculated evidence of insufficient time, any significant decision arising from the computer's application of the controller's policy, will probably have to be presented to the controller for approval—although some of the reasons for this may be more psychological than logical. 'Significant decisions', in this context, would probably be those manoeuvres involving less than the normal safety margins and those involving changes in order of arrival or departure, deviation from schedule beyond normal operating tolerances, major diversion of route, and change of destination.

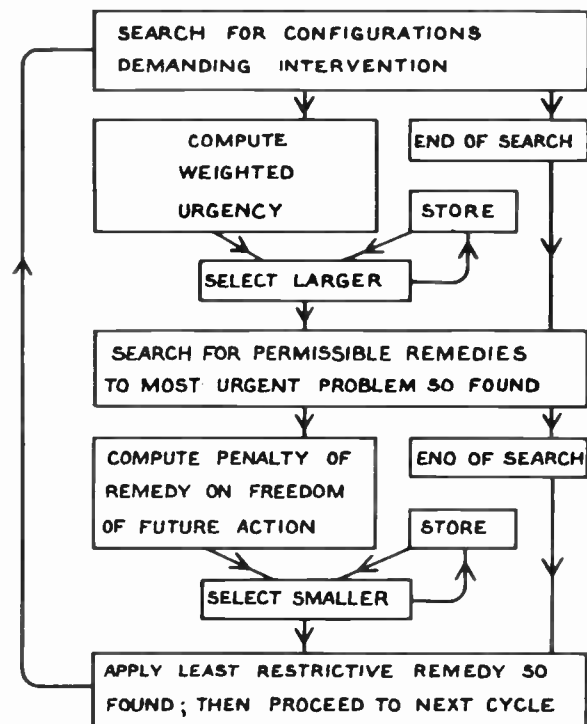


Fig. 1. Operational control routine.

3.3. The Basis of a Possible Computer Program

In a complex situation, it may strain even the capacity of an electronic computer to calculate rigorously that combination of control orders which will optimize the suitably-defined joint performance of all relevant vehicles. However, an adequate approximation to this ideal might be obtained by setting a threshold defining the degradation of performance or safety, beyond which intervention by the

control system is warranted. The basic strategy of the computer program might then be developed along the following lines:

- (i) To search for 'situations' warranting intervention.
- (ii) To arrange these in a chronological order of urgency.
- (iii) If necessary apply some bias to the urgency according to the classification of the 'seriousness' of the situation.
- (iv) For the most 'urgent' situation not yet dealt with, find those adequate remedies which are unlikely to create further relevant 'situations'.
- (v) In particular avoid remedies which might create or affect situations of greater 'urgency' rating.
- (vi) Of the permissible remedies thus found, adopt that which least restricts the system's 'freedom of movement' in coping with other situations which have not yet been considered—or indeed have not yet arisen.
- (vii) Proceed to the most 'urgent' situation then remaining.

This routine is shown schematically in Fig. 1.

3.4. Potential Military Applications

Although the problem so far has been primarily considered in terms of air or possibly sea traffic control in terminal areas, it may apply even more critically in a military situation.

In the stress of battle, a Commander may be so busy allocating specific weapons or units to cope with specific enemy units or targets, that he can give little thought to his overall tactical aims or to the deployment or replenishment of his reserves. This may restrict his ability to vary his operational doctrine as a function of time or sector, or type of enemy unit or target, or to assess the impact of earlier engagements on his subsequent tactical action. Thus, the heavy load of routine, 'machine-like' tasks might leave a Commander with insufficient margin to adjust his operational doctrine to those subtle tactical developments and patterns which a skilled Staff Officer could otherwise appreciate, but which at present defy any generalized formulation suitable for automatic recognition.

Hence, computers may be of value both to assist in the executive application of the established tactical doctrine and for the compilation of a comprehensible, up-to-date, overall picture, on which the battle Commander can base his tactical decisions. With a purely manually-assisted command structure, some vital decisions might have to be delegated to men at a lower level, who are aware of only part, even of

those relevant data arising within their own area or sector.

3.5. Inputs to an Automatic Control System

An automated traffic surveillance and control system is shown in schematic form in Fig. 2. In such a system, virtually all information on the position and movements of non-co-operating vehicles and on the position, movements, capabilities, current assignments and orders of co-operating vehicles would be contained in a common data store. Most of this information would arrive from a variety of sensors and data-links, automatically and/or by manual injection from a number of specialized operators. However, there is a considerable body of (normally) longer-term information which could be inserted directly into the store. This would cover such items as:

- the capabilities of various classes of vehicles,
- operating schedules,
- zones of responsibility,

as well as operational procedures, doctrines and policies. Modifications to the general operational policy and individual exceptions to it would also be inserted directly into the machine.

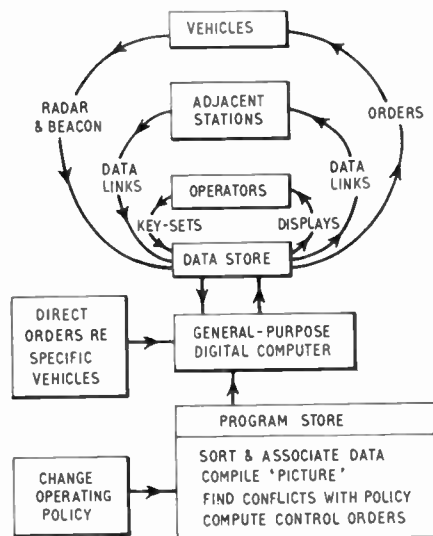


Fig. 2. Schematic diagram of automatic control system.

Hence, considerable effort must be devoted to devising suitable message formats for the various data links which transmit status information, procedural instructions to regulate the information flow itself, and operational orders. A convenient, flexible, standardized format and control panel will also be required for the manual insertion, into the machine, of information, policies and decisions (and requests for specific information). Technologically, it would

be possible—though expensive—for a computer system to accept all these incoming data in a generous vocabulary, allowing for a wide choice of possible forms of expression. However, it is both more economical and a helpful discipline and *aide mémoire* for the user to put his information into a standard format. The format of information inputs and outputs should, however, be designed primarily for the convenience of the user. The machine can then perform any necessary interpretation or translation.

3.6. Filtered Information Outputs

Having seen what sort of information is likely to be available, we can consider its distillation, to match it to the man:

(i) Where essentially the same information comes from several sources, the machine can obviously combine these data. Where appropriate, it can take a weighted mean of the incoming data, and possibly assess a suitable composite 'confidence' figure for the resultant estimate.

(ii) Where a new report conforms to an already known movement, the machine can confirm and up-date the previously extrapolated position (and velocity).

(iii) Where a new development conflicts with the criteria of desirability, success or safety associated with the operational doctrine in force, this situation must be indicated to the appropriate computer program and/or controller.

(iv) Where the manoeuvres of a number of individual vehicles constitute—or contribute to—a common systematic pattern, this could in principle be recognized automatically and then presented to the machine or man as a single, more complex, development:

For instance, the recognition of systematic movements over a given traffic lane, terminal area or zonal boundary might well become a machine function. This would also assist in maintaining the track identity of individual elements within this pattern of movement. However, in its more general form, the recognition and evaluation of patterns in the operational 'picture' seems better suited to the human brain than to any likely machine.

(v) When new information concerns only vehicles at present beyond a station's zone of responsibility, the relevant items can automatically be brought up-to-date in the data store without reference to any operator.

3.7. Matching the Machine to the Man

3.7.1. Facilities required by a controller

A good deal of study has been devoted to the formulation of operational doctrines as a set of rules in

a computer. Our aim must clearly be to make these rules sufficiently comprehensive to express any foreseeable policy or doctrine by appropriate adjustments of a set of 'weighting parameters'. At the same time, these weighting parameters must be sufficiently few, mutually independent, and simple in function to permit them to be quickly and effectively adjusted in response to any critical new development.

Much thought has also been given to the requirements for tabular information and symbolic plan displays to produce operational control consoles with versatile displays and injection facilities. These consoles should be equally suitable for the control of over-all policy or, when the machine computes that sufficient 'thinking time' is available, for the monitoring and possibly 'human modulation' of the machine's proposed orders to individual units—or, in the military case, weapons. The computer-generated orders should, of course, faithfully reflect the operational rules and doctrines previously inserted into the machine and perhaps subsequently modified in the light of later developments.

3.7.2. Displays

Clearly, each operator should be matched to the remainder of the control system in the optimum manner for his specific task. Hence, each man should probably be provided with his own console, comprising a plan display, a tabular display, a computer-injection panel and a control panel for conventional communications. With computer control of the particular synthetic plan and tabular displays currently required at each operating position (and possibly even of the communications facilities), all these consoles can be of identical design.

Although it would be most inadvisable not to provide the option of 'raw' radar display, it seems likely that a user would normally choose a synthetic plan display of those target categories and symbols, and those range and height limits relevant to his current job. Computer-generated symbols should appear quasi-continuously on such a display, but selected markers might be made to flicker distinctively, to draw attention to themselves. Colour might be used as a valuable additional means of distinguishing one symbol from another, but there is much doubt as to the wisdom of relying on colour exclusively for indicating a vital distinction between display symbols. For air operations, afterglow persistence is likely to give a satisfactory visual indication of courses and speeds. For surface navigation, dot markers, associated with the symbols, might be made to repeat the target's movements, over the last operationally significant period, repetitively, on a suitably accelerated time-scale.

The tabular display might conveniently be divided

into several sections. For each of these, there could then be an independent choice of a number of alternative sectional displays. Each user could thus choose independently which (if any) of the available alternatives he wishes to see on each section of his display. In practice he would probably select one of a number of predetermined combinations of sectional displays.

Full facilities would be required for cross-reference between the tabular and plan displays, and for cross-reference between either form of display and the data-processing computer, when injecting or demanding additional information.

3.7.3. Conference facilities

The individual pictures would of course be derived from a common body of digital information, and the users could have both versatile voice inter-communications, and facilities to display common electronic markers and joystick-controlled 'pointers'. These provisions should permit any combination of local or remote users to combine in a 'conference' net, without leaving their own posts, displays, control panels, communications, etc. It is thought that conference requirements may in the future be increasingly met along these lines, possibly supplemented by television or facsimile, if and where necessary. In other words, it may prove preferable to bring the *relevant* information to its users, rather than bringing all the users to *all* the information, using a single, common, outsize display.

Only visual inputs to the human being have been discussed above. Warning buzzes might also be desirable, in certain circumstances, and pre-recorded verbal messages could evidently be used, if required. However, it is thought that most operational information is more appropriate to a persistent, parallel, geometrical, visual representation than to the transient, serial, verbal, auditory form. Furthermore, it is judged to be both convenient and psychologically desirable to reserve the auditory input mainly for conventional voice communication and normal human contact, and for the 'conference' facilities described above.

3.7.4. Problems of continuity in the introduction of automation

The engineer's ideal is naturally a flexibly deployed integrated system of men and machines, whose capacity is just matched to the maximum likely overall load. By definition, such a system would keep to a minimum the size and complexity of the human organization within the system, the computer programs and the actual 'equipment'. It should result also in the scheme that is operationally and technically least difficult to understand—for the newcomer.

However, such an idealization leaves out of account

that the system will be manned by existing operators. Hence we must pay some regard to the hierarchical structure, allocation of responsibilities and general operational background arising from their training and experience. Indeed, the development of suitable tactical rules, display facilities, operational doctrines and procedures, indoctrination and training programmes, etc., merits as much attention as the sophisticated equipment, which has to go with these to make up an effective system of automated extraction and partially automated utilization of radar information.

3.7.5. Simulation facilities

Systems based on stored-program computers have inherently very full and realistic capabilities of simulating their own intended operation with synthetic, self-generated inputs, recording their own operation, analysing their own performance.

This facility of self-simulation, recording and analysis could play an invaluable part in:

- the initial training of the operators,
- subsequently maintaining their skill and interest,
- extending, refining—or revising—the capabilities of the system,
- adapting the system to new requirements, and general operational studies and assessments.

These facilities should also be of great value in testing such systems, both during development and when in operational use, as regards the suitability and correct functioning of:

- (a) the computer programs defining display configurations and facilities,
- (b) the programs defining the manual-injection arrangements,
- (c) the programs defining the operational rules,
- (d) the actual technical equipment, insofar as it is of computer type or under computer control.

3.8. Reliability in Operation

3.8.1. The role of the maintainer

To prevent the automatic element of a traffic control organization from 'going mad', some form of automatic fault indication is vital. When the existence of a fault is thus indicated, automatic diagnosis can speed up the logical search for the faulty element by a factor of, say, 10^6 . The subsequent manual insertion of a plug-in spare element would then still be extremely rapid, compared with the time taken by human diagnosis.

Human maintainers are likely to remain responsible for fault finding in relatively simple peripheral

equipments, and for the operation of specialized test logic, to establish the correct operation of the minimum of computer equipment to permit the auto-diagnostic programs to function. With time-shared, stored-program computers, the special test gear and auto-diagnostic programs need not constitute a very significant addition to the cost and complexity of the overall system.

Compared with non-automatic equipments, both the required number and the calibre of the routine maintainers is likely to be reduced. This arises from the small number of binary logical circuit designs required, from their relative simplicity and high reliability, and from automatic or semi-automatic fault indication and diagnosis. However, when difficulties arise which defy the automatic tests, maintainers of very high calibre would be required. New problems may arise in the selection of these men, their training, and the maintenance of their—in-frequently exercised—skill and familiarity with the system.

3.8.2. Design considerations

Where much of the executive aspects of operational control is delegated to an automatic computer, it is vital that this computer should both inspire and deserve the utmost confidence. A great deal of attention must therefore be given to the storage of the operational rules and policy in a manner which permits them to be altered readily, when required, but which is immune from accidental change due to any electrical malfunctioning and has a good resistance to mechanical damage. Furthermore, the computer program must be protected with extensive error-detection and 'fail-safe' facilities.

Any installation of this type must have very full electronic self-monitoring and auto-diagnostic facilities. Moreover, it must have sufficient redundancy in its basic equipment that any likely defect not covered by these diagnostic facilities would still permit the system to function—though at reduced capacity or sophistication—rather than causing a complete breakdown. For instance, the various sensing devices at each station should be fully integrated, to reduce the dependence on a single sensor (or its automatic data-extraction equipment).

Finally, the individual stations making up a control network should interchange all their information by means of appropriate digital data links, so that each station holds the aggregate information derived from all their joint sources of information. This redundancy ensures reasonable continuity of the overall operation, even if any station (or its primary data source) is out of action, or if it loses communications contact with its neighbours.

4. Conclusion

Experience from research on automatic radar data extraction provides a reasonable basis for the prognostication that the application of similar principles to the automation of many operational functions is both feasible and worthwhile. The further problems to be solved, in applying the results of this research to any specific operational application, are as much in the fields of procedure, organization and training as in science and technology, and their solution depends on the creation of a fully integrated 'scientific-user' team.

Manuscript received by the Institution on 1st April 1963. (Paper No. 854/RNA20).

© The British Institution of Radio Engineers, 1963

(The discussion on this paper is printed on pages 327 and 328.)

Automatic Radar Data Extraction by Storage Tube and Delay Line Techniques

By

J. C. PLOWMAN, M.Sc.†

Presented at the Symposium on "Processing and Display of Radar Data" in London on 16th May 1963.

Summary: The basic recognition criteria employed in the extraction of radar target data are described and the application of these criteria in human and automatic radar data extraction systems is discussed. The details of two experimental auto-extraction systems are given. One system involves the use of storage tubes and the other employs fused quartz delay lines as the necessary storage medium. Qualitative results with both systems are discussed.

1. Introduction

The modern surveillance radar system is required to give a rapid and accurate assessment of target position in a world of decreasing target echoing areas, increasing target speeds and often in an environment of severe clutter or interference. Further, it may be required to give this assessment as yet another contribution to data handling systems which are continually dealing with information from a number of local and remote sources. The volume of data available and the necessary speed and accuracy in handling suggests the solution of automatic handling of data by digital computers.

Clearly the data extraction system associated with any data source, be it radar or other, should match the potential of that source to the capacity of the overall data handling system. The human operator/radar display combination fails to do this efficiently in a radar/computer environment and thus methods for the automatic recognition of radar signals are being investigated.

This paper will deal with the basic recognition criteria applied to radar signals, discuss operator failings, and outline the salient features of two auto extraction systems investigated for use with pulsed radars.

2. Radar Signals

Figure 1(a) shows a radar aerial rotating clockwise about an axis through O. As the radiation pattern or 'beam' sweeps across a point target T the returned signal from T will increase to a maximum and decrease as shown in Fig. 1(b).

The vertical lines of Fig. 1(b) represent the ideal returned pulse amplitudes from successive transmission

pulses occurring as the beam sweeps across the target position. The maximum amplitude of the group of returned signals is a function of the transmitted power and of the range and echoing area of the target.

The number of returns which effectively contribute to a group of signals returned from a point target depends on the beamwidth and rotation rate of the aerial and on the pulse repetition frequency (p.r.f.) of the radar transmitter.

The amplitude relationship of returns within the group, i.e. the 'shape' of the group, is a function of the two-way radiation pattern of the aerial. When the amplitudes are considered on a logarithmic scale then the 'shape' of this amplitude envelope remains constant about its peak, irrespective of the maximum amplitude and assuming that amplitude fluctuations due to system noise, target glint, etc., are neglected and that a point target only is considered. Targets

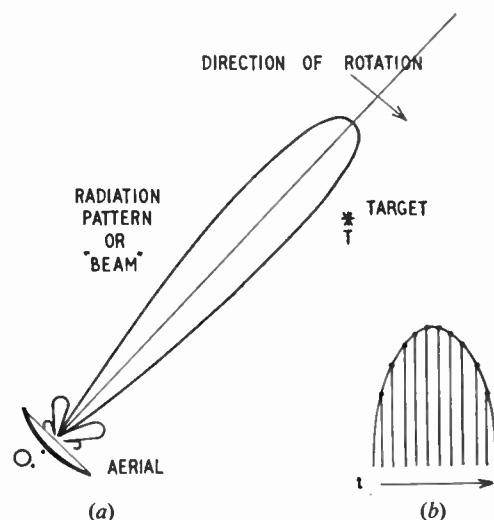


Fig. 1. The generation of a group of radar returns from a point source target.

† Admiralty Surface Weapons Establishment, Portsmouth, Hampshire.

which are large compared with the projected area of the beam will not provide this shape of amplitude envelope.

The duration of each return from a point target is determined by the duration of the transmitted pulse and to some extent by the bandwidth of the radar receiver. Returns from targets elongated in range—e.g. land returns may be of a duration greater than that of the transmitted pulse.

3. Recognition Criteria

From the considerations outlined above the following criteria for the recognition of point targets can be laid down:

- (i) The mean energy of a group of signals shall exceed a threshold.
- (ii) The 'shape' of the group of signals shall conform with that of the known two-way radiation pattern of the aerial within a given tolerance.
- (iii) The 'duration' of pulses in the group shall be less than a predetermined maximum defined by the duration of radar transmitter pulses.

4. Basic Data Extraction System

In order that the above recognition criteria for amplitude, shape and duration can be applied it is first necessary to collect a beamwidth's worth of radar returns, which are available in serial form, and to present the resultant groups of returns as a whole for recognition. Hence the essential parts of a radar data extraction system are as shown in Fig. 2; that is, a store inspected by a recognition device.

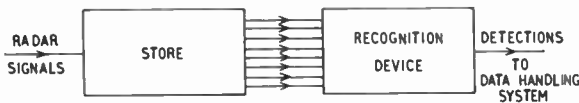


Fig. 2. A radar data extraction system—basic configuration.

The store must provide storage for signals contained in at least one beamwidth over the whole range bracket concerned, and further, it must correlate returns from the same range elements in successive transmission periods and present the resultant groups of signals, range element by range element, to the recognition device.

5. The P.P.I. Display and Human Operator

This combination has long been a conventional data extraction system employed with radars. It can readily be discussed in terms of the recognition criteria and basic data extraction system mentioned above. A typical p.p.i. or 'Panoramic' display of radar data is shown in Fig. 3. This picture is obtained

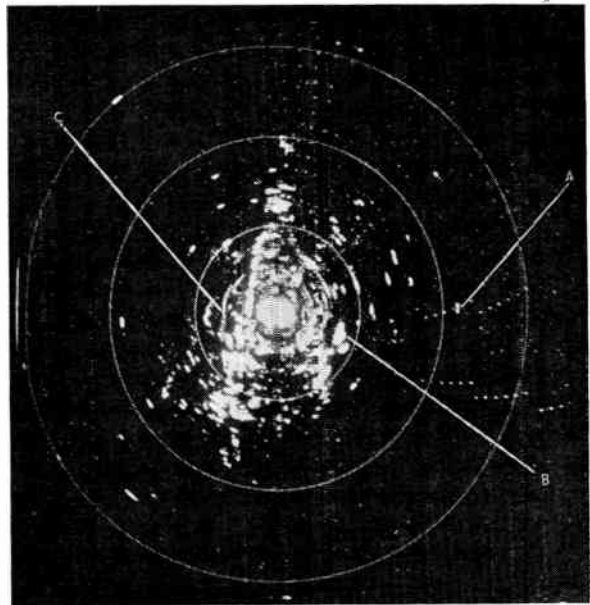


Fig. 3. A typical p.p.i. radar picture (1 revolution).

on the long after-glow screen of a cathode ray tube (c.r.t.) by the use of suitable electronic deflection and 'bright up' circuits.

In terms of the basic system, the necessary storage medium is provided by the afterglow phosphor of the c.r.t. screen. The deflection and bright-up circuits 'paint' the radar information on to this storage medium in a manner which ensures that data from corresponding range elements, in successive transmission periods, are presented in adjacent storage areas and the plan picture so obtained presents the serially available radar input data to the observer in a coherent 'parallel' form.

The storage capacity of the p.p.i. display is, in general, determined by the spot size and screen area (i.e. by c.r.t. definition) and this capacity can amount to 10^6 storage elements—sufficient capacity for the temporary storage of point source targets in any, or all, of about 10^5 positions. This represents a considerable amount of storage, all positions of which must be inspected by the operator.

The human operator is of course the recognition device in this data extraction combination. The operator compares the information presented by the p.p.i. picture with memory patterns developed from past experience or from deduction. Briefly, these memory patterns demand a bright arc of a certain length and of a certain radial thickness, if a point source target is to be recognized, and its range and bearing assessed and passed to the data handling organization. Thus in the recognition process:

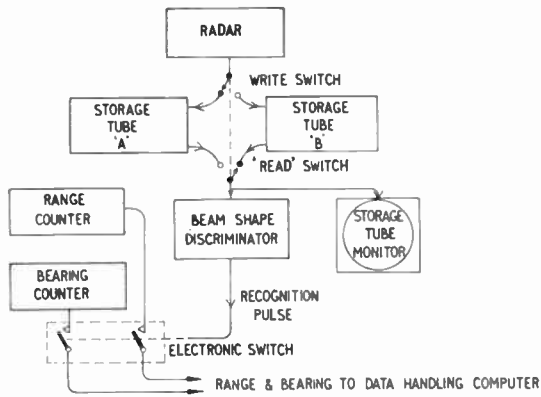


Fig. 4. Storage tube data extraction system—block diagram.

- (i) a threshold is exceeded in that the c.r.t. screen is brightened,
 - (ii) a rough beam shape must be achieved in that the arc length is to be between required limits, and
 - (iii) the expected pulse duration must not be exceeded.
- Hence the three required recognition criteria as laid down are imposed by the human recognition device and, for example, in Fig. 3, A is identified as a point source target but B and C are not.

It should be noted that the radar cathode-ray tube screen is virtually an on/off device and thus all data is assessed against a fixed background level with an efficiency determined by operator temperament, training and fatigue. Shape information is crude and hence bearing accuracy can suffer. Spot size, range scale and deflection circuits influence the range accuracies achievable.

6. Operator Failings

The human operator fails as an efficient and economical extractor of radar data due, in the main, to his slow speed of reaction and assessment, his poor accuracy, rapid saturation and his lack of long-term concentration. These failings can, to some extent, be combatted by limiting the sector to be inspected and by the provision of 'joystick' controlled markers interlaced with radar presentation.

However, in spite of these and other facilities—resulting in the ingenious and complex display systems associated with powerful radars—it has been shown that the human operator cannot 'up date' tracks at a rate much greater than once per two seconds and also that his efficiency in the initial detection of new targets decreases rapidly after the first half hour or so. These factors would result in almost prohibitive manpower requirements if the full potential of modern radars were to be matched to the capacity of a data handling system by human operator/radar display combinations.

7. A Storage Tube Auto-Extraction System

The storage tube was initially chosen as a currently available storage medium for research purposes, and an automatic detection system was built employing cathode potential stabilized Emitrons, with a simple recognition device.

7.1. System Diagram

A block diagram of the storage tube systems is shown in Fig. 4. Automatic detection is performed by the use of two storage tubes: the received radar information from consecutive bearings sectors is written on to these tubes in turn. Due to the resolution of the storage tubes information from only a limited range bracket (e.g. 60 to 120 miles) and a limited bearing sector of 15 deg could be stored on each tube.

7.1.1. Write cycle

In a given 15 deg sector of aerial rotation the received radar information (logarithmic video) is 'written' into one storage tube, say tube A in Fig. 4, as a range-azimuth pattern, Fig. 5(a), i.e. range scan vertically and bearing displacement horizontally. The storage tube mosaic then holds a charge pattern corresponding to the received radar data until the tube is 'read' in the next 15 deg of aerial rotation; tube B is then storing the incoming data. Thus each tube records alternative 15 deg sectors.

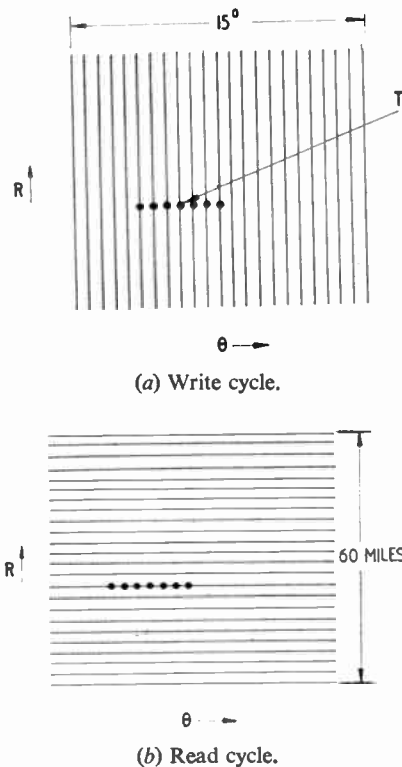


Fig. 5. Storage tube system—scanning patterns.

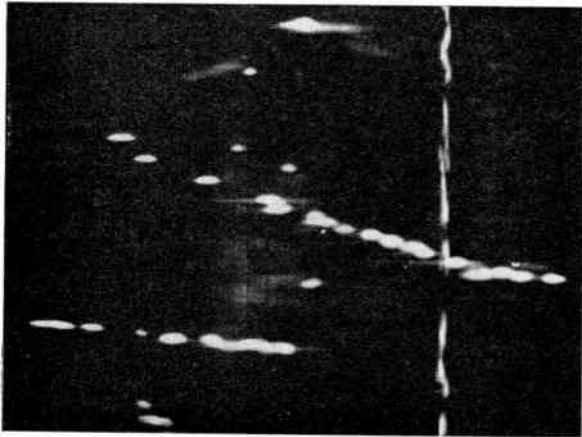


Fig. 6. Storage tube charge pattern displayed on a monitoring tube during the read cycles. The horizontal streaks are the read out signals and the bright spots are the superimposed auto-detection markers.

7.1.2. Read cycle

During a read cycle, the storage tube mosaic is scanned horizontally in the bearing direction and the start of each scan is stepped vertically in the range direction in unit increments of range (Fig. 5(b)). Thus a charge pattern, such as at T, formed by successive returns in the write period, produces an output when the read scan passes through its position. This output is a pulse with a shape which depends on the beam shape of the radar aerial.

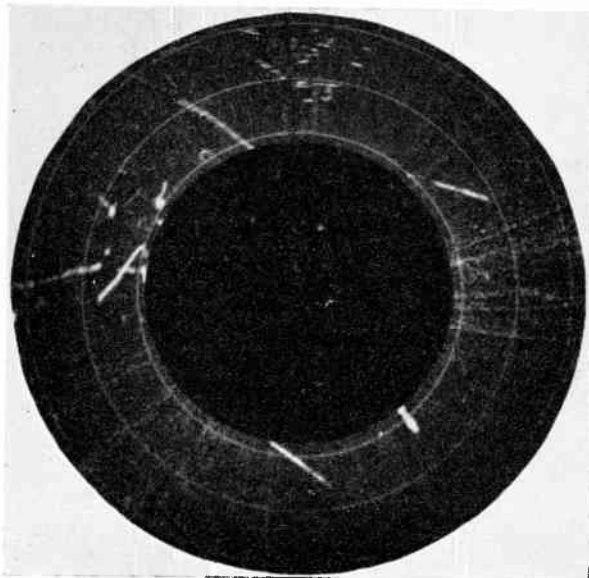
7.2. Echo Recognition

The output from the storage tube during the read cycle is passed into a short lumped constant delay line. The delay of this line is determined from the scan speed in the read cycle and on the beamwidth of the radar involved. At any instant therefore the envelope of a beamwidth's worth of radar information, corresponding to a given range element, will be distributed along the delay network. If the reading scan happens to be scanning through a recorded target, this envelope will be similar in shape to the aerial two-way radiation pattern.

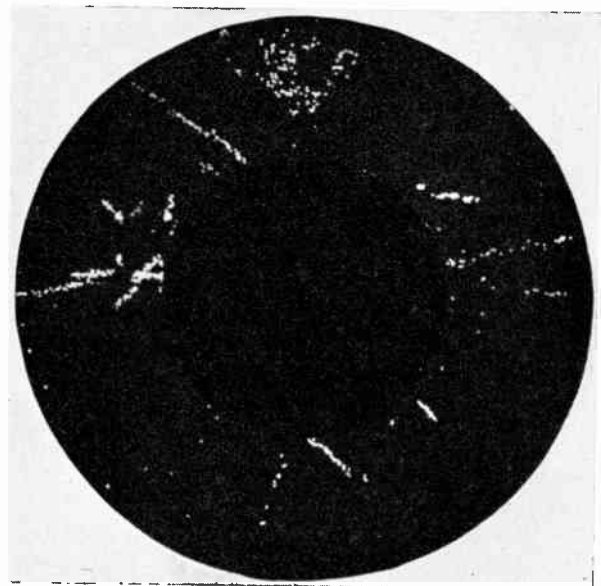
The outputs from seven taps, situated at equal intervals along the delay line are routed to a network of comparator circuits which determine when the envelope is greater than a given threshold, is centred in the delay line and has a predetermined shape. When the amplitude and shape are acceptable a recognition pulse is generated which gates range and bearing data to the data handling computer. (Range and bearing are available in digital form from counters which are synchronized to the deflection waveforms used in the read cycle.) No 'duration' criteria was imposed in the experimental storage tube system.

7.3. Results

The system produced results which demonstrated the real potential of auto-detection associated with a suitable digital computer. These results are illustrated by Figs. 6 and 7.



(a)



(b)

Fig. 7.

(a) Raw radar p.p.i. (10 minute exposure).

(b) The corresponding auto-detected positional display produced by the storage tube auto-extraction system with subsequent processing by an associated digital computer. Correlation between radar and synthetic tracks will be noted.

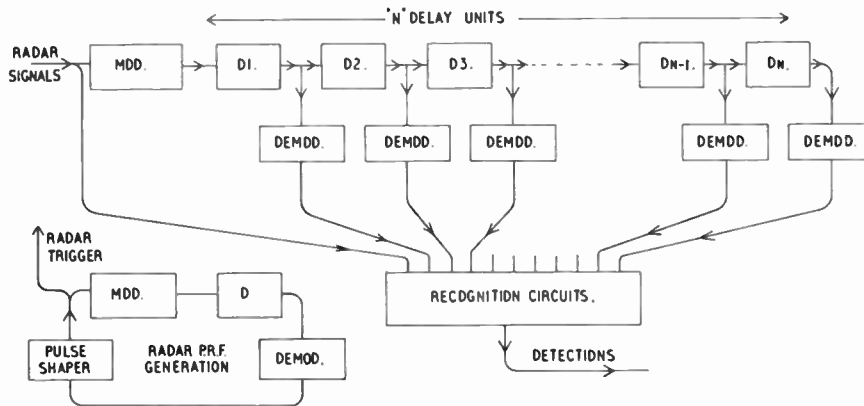


Fig. 8. Delay line auto-extraction system—block diagram.

Figure 6 shows, on a monitoring oscilloscope, the charge pattern read off the storage tube together with superimposed auto-detection markers. The bright spots are the markers and the horizontal streaks are the storage tube read out signals. The exposure corresponds to 10 or 12 aerial revolutions and the tracks of identified aircraft are clearly seen.

Figure 7(a) shows a raw radar p.p.i. picture (10 minute exposure). The corresponding output of the automatic detection system (after processing by the associated digital computer and then suitably converted for display) is shown in Fig. 7(b). Correlation between the radar tracks and those produced by the auto-extraction system is readily identified.

8. Delay Line System

An alternative approach to the 'static' storage of data by storage tubes is that of 'dynamic' storage as provided by delay lines.

The delay line introduces a predetermined time delay, t , between its input and output terminals. Hence data entering the delay line can be compared with that emerging from it, and thus input data entering can be compared with data inserted time t earlier. By employing a chain of n equal delay circuits, dynamic storage of radar data from the n previous transmissions can be achieved and continual correlation of fresh data with that obtained in the previous n radar periods is possible.

It will be recalled that the number of 'effective' radar returns from a point source target depends on the radiation pattern, aerial rotation rate and p.r.f. of the radar. The number of 'effective' returns is defined as that number required to present a reasonable shape for the group. If the radar parameters are such that the period between transmissions is t and the number of 'effective' returns is $N+1$ then a chain of N delay circuits, each of delay t , will provide the

necessary storage. A system employing this principle is shown in Fig. 8.

Signals from the radar receiver modulate a carrier frequency which is injected into the input of the chain of delay units—the signals pass down the chain from delay unit to delay unit.

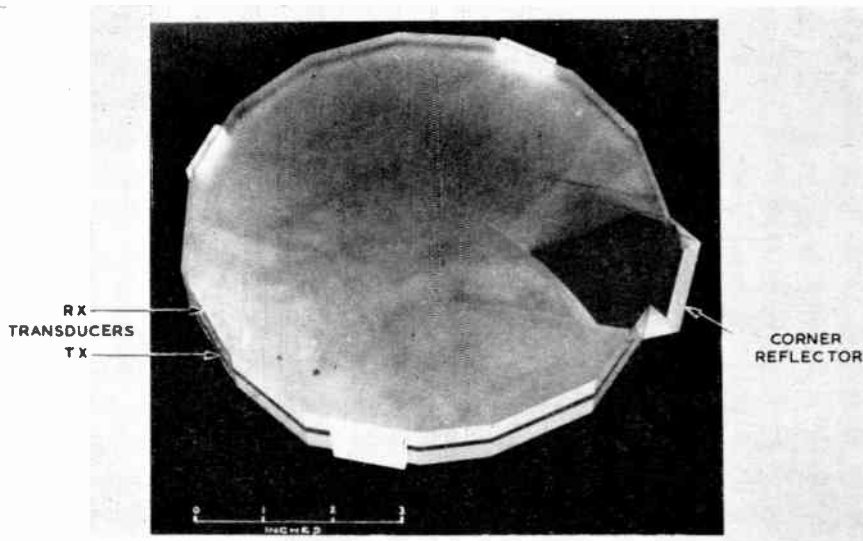
To ensure that the period between radar transmission pulses is constant and equal to the delay, t , of each of the delay units in the chain, the radar trigger pulses are generated from a recirculating loop which includes a similar delay unit.

Since the delay of each unit in the chain is exactly equal to the period between radar transmissions it will be seen that at any given transmission, radar data from N previous transmissions are held in the delay chain. Thus at any instant, after a given transmission, the demodulated output signals appearing at the tapping points between delay units in the chain correspond to the radar returns in the same range element in the N previous radar periods. Hence a beamwidth's worth of returns for each consecutive range element can be simultaneously presented to the recognition devices.

8.1. The Delay Unit

8.1.1. Requirements

The use of a chain of delay units necessarily implies very accurate delay timing in each unit since, for example, time coincidence (within $\pm 0.1 \mu s$) of a $1 \mu s$ pulse at the n th tap with a pulse of similar duration at the input is necessary. The delay time from input to N th output can be say 10 milliseconds or more and thus a delay accuracy of at least 1 in 10^5 is required. Further it is also necessary in each delay unit to maintain an accurate gain of unity throughout the dynamic amplitude range. This ensures that there is no unacceptable build-up, or decline, of amplitude as a signal passes along the chain.



(a) Line configuration.

(b) Constant temperature housing.

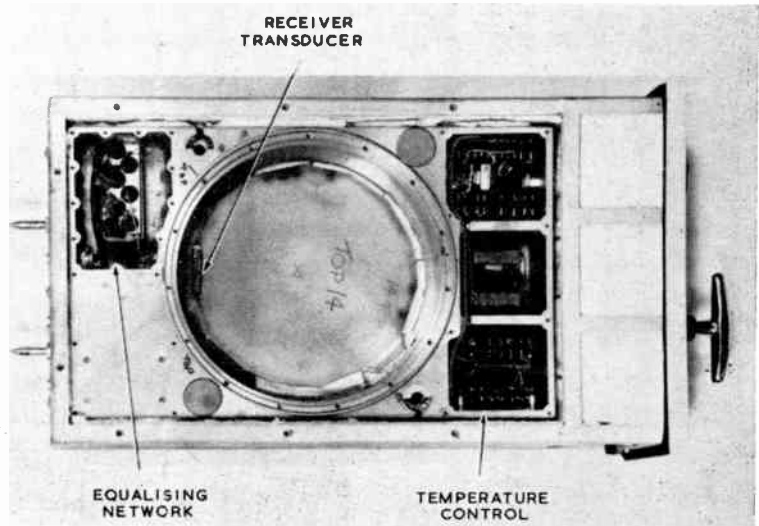


Fig. 9. Delay line unit.

8.1.2. Configuration

These considerations lead to the choice of ultrasonic fused quartz delay lines as the delay elements. A delay of roughly 2 ms can be obtained by a 'double decker' fifteen-sided polygon construction as illustrated in Fig. 9(a)—all sides are accurately ground to ± 3 seconds of arc, and the upper and lower 'decks' are connected by a corner reflector.

Ultrasonic waves, with a carrier frequency of about 30 Mc/s, are launched in the shear mode by quartz crystal transducers bonded to a given face on one deck of the polygon. The waves traverse an elaborate zigzag path by multiple reflections within the quartz block until they are eventually detected by a receiver

transducer cemented to the same face on the other deck of the block.

8.1.3. Delay stability

The delay of a line depends to some extent on temperature. The type of line described has a delay time/temperature coefficient of approximately 1 in 10^4 per deg C and hence each delay line is individually temperature controlled to ± 0.1 deg C. This control is achieved by maintaining each line at about 10 deg C above a common oven temperature of about 55 deg C. The individual temperature control is used to effect small adjustments to delay time and so correct the manufacturing tolerances. A delay line in its housing is shown in Fig. 9(b). All delay lines including that in the p.r.f. loop are housed into a common oven cabinet.

8.1.4. Gain stability

Each delay line introduces a loss of about 60 dB—due in the main to transducer losses. This loss is regained in a wide band amplifier which is associated with each line. The output of this amplifier feeds the next line in the chain. A delay line—amplifier combination is referred to as a delay unit.

Gain stability is achieved by automatic gain control and the tight linearity specification is met by suitable amplifier design.

The bandwidth of a delay unit (including amplifier) is approximately 10 Mc/s about a centre frequency of 30 Mc/s, and a flat response in the pass-band is achieved by individual equalization circuits associated with each delay line.

8.2. Application of Recognition Criteria

The recognition criteria of amplitude, shape and duration as described earlier are continually applied to the group of $N + 1$ outputs simultaneously presented to the recognition circuits by the demodulators attached to the tapping points in the delay chain.

8.2.1. Amplitude

It will be recalled that with the p.p.i. display the amplitude of a signal group was assessed against a fixed background due to the on/off characteristics of the display. Delay line storage however permits the weighted amplitude within a signal group to be assessed against the mean background level of its immediate environment.

This can be achieved as shown in Fig. 10(a) by arranging that the outputs (ω) from all the taps are added together thus obtaining the mean amplitude,

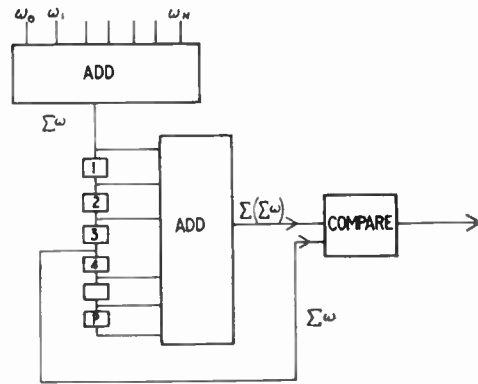
$$\left(\frac{1}{N+1} \sum_{n=0}^N \omega_n \right)$$

This resultant amplitude is now passed down a chain of P short lumped constant (LC) delay lines where each delay in this chain is equivalent to one range element. This delay chain will at any instant contain the *mean* amplitudes obtained from P consecutive range elements. The mean of the sum of the outputs at all the taps in this chain,

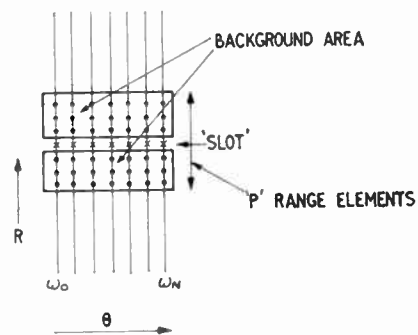
$$\left(\frac{1}{P+1} \sum_{r=0}^P \left(\frac{1}{N+1} \sum_{n=0}^N \omega_n \right)_r \right)$$

provides a measure of the mean amplitude in an area of $(N+1)$ radar periods (in bearing) by $(P+1)$ elements in range.

Now if the centre tap of the chain of P lines is not connected to the range summation circuit then the output of this circuit is a measure of the mean ampli-



(a) Block diagram.



(b) Explanatory diagram.

Fig. 10. Signal amplitude and background comparison.

tude of the area surrounding the range 'slot' represented by the centre tap, as is shown in Fig. 10(b). Signals in the 'slot' do not contribute to the background measurement and the output from the centre tap is a measure of the mean amplitude of the signal group in the slot—and thus if this output is compared with that of the range summation circuit then the mean signal amplitude is being compared with the mean of the local background, i.e. with a floating threshold, which must be exceeded by a preset amount to meet the amplitude criteria.

The above simple description serves to indicate the principle of the floating threshold. In a practical realization of this concept several sophistications might be employed such as optional emphasis of the background and weighting of the individual values in the group for determination of the weighted amplitude in the 'slot', etc.

8.2.2. Beam shape discrimination

Simultaneously with the amplitude assessment described above the 'shape' of the incoming group of signals from the delay line demodulators is compared with a voltage 'template' representing the two-way radiation pattern of the aerial.

It is necessary to use a logarithmic radar receiver when beam shape is to be recognized since this ensures that the shape about the peak is independent of the mean amplitude of the group of target returns.

When an incoming group of signals is of perfect shape and is centred in the delay line chain, then the only difference between this group and the corresponding template voltages will be due to the difference in relative mean amplitude, since the template represents a fixed amplitude signal, corresponding to x dB about the tip of the beam. Thus, to obtain a zero error between a perfect incoming group and the template, it is necessary to "normalize" the incoming group (by subtracting its mean amplitude) to the same level as the template before determining the error.

Since the template is arranged to correspond to a perfect shape centred in the delay chain then, when a 'perfect' group moves down the delay chain, the resultant error will decrease to zero as the group shape passes through the centre and will then increase again. Thus, although the best 'fit' is obtained with the signal group central in the delay chain, fairly good fit indications can occur one radar period earlier or later and precautions must be taken to reduce the generation of multiple 'fits' to a minimum.

8.2.3. Beam shape discriminator circuits

A block diagram, shown in Fig. 11, illustrates a method of beam shape discrimination designed to meet the above requirements. The template is generated as a series of negative voltages $-\lambda_0$ to $-\lambda_N$. These voltages are permanently connected to a series of analogue adding networks which receive their inputs, ω_0 to ω_N , from the demodulators associated with the delay chain. Thus in this series of adding circuits λ_i is subtracted from the signal input, ω_i , from the i th

tap of the delay chain. The resultant outputs $(\omega_i - \lambda_i)$ are now summed and a mean taken to give a measure of the mean amplitude of the input group and hence to provide the 'normalizing' voltage:

$$V_N = - \frac{1}{N+1} \sum_{i=0}^N (\omega_i - \lambda_i)$$

In a second series of analogue adders this negative normalizing voltage is added to each $(\omega - \lambda)$ term and thus the output from each of these adders gives a measure of the error between the input amplitude concerned and the corresponding template value. This error can be positive or negative and hence the output of the j th second adder is given by

$$\epsilon_j = K_j \left(\omega_j - \lambda_j - \frac{1}{N+1} \sum_{i=0}^N (\omega_i - \lambda_i) \right)$$

where K_j is a weighting factor dependent on the slope of the input/output characteristic of the logarithmic receiver at points corresponding to λ_j in the template. This factor partially linearizes the individual errors and hence reduces the significance of errors resulting from the outer constituents in the group of signals. The factor assists in the detection of weak signals in noise.

The resultant instantaneous total error in shape of the input signal group is given by the sum of the modulus of all the constituent errors, i.e.

$$E = \sum_{j=0}^N \left| K_j \left[\omega_j - \lambda_j - \frac{1}{N+1} \sum_{i=0}^N (\omega_i - \lambda_i) \right] \right|$$

This result is provided by the use of two summation amplifiers as shown.

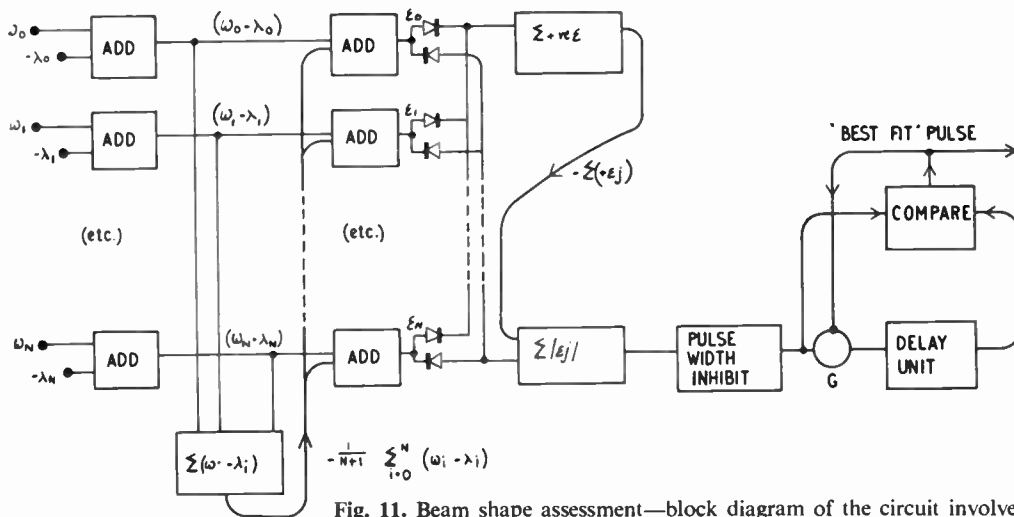


Fig. 11. Beam shape assessment—block diagram of the circuit involved.

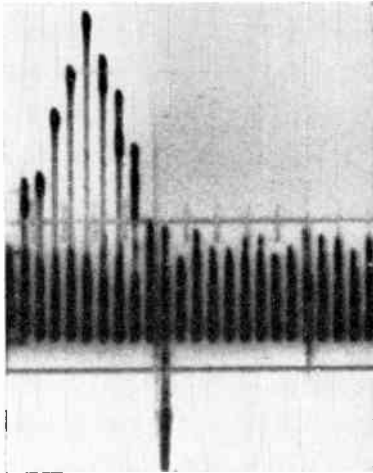


Fig. 12. A typical group of returns from an aircraft (logarithmic radar video with recognition pulse superimposed).

The analogue computer which evaluates the above expressions is in practice a simple combination of d.c. restoring and resistor networks associated with three multi-input see-saw amplifiers. The whole device can be accommodated on two printed circuit cards each about 10 in by 7 in (25 cm × 18 cm) if transistor circuit techniques are employed throughout.

The circuit arrangements are such that a perfect fit, i.e. zero error, gives an output of -5 V. A poor fit, due to receiver noise inputs, is the normal state of affairs and thus a good fit is indicated by a negative-going pulse which exceeds a normal level of about -2 V. This negative-going pulse is known as the 'fit' pulse and it must exceed a preset threshold if it is to be accepted as a potential recognition.

8.2.4. Pulse duration

The application of this criterion will not be described in detail and it will suffice to say that the duration is checked by means of a short preset delay circuit which inhibits further action if the pulse duration is greater than expected.

8.2.5. Multiple fit rejection

The amplitude of a 'fit' pulse is a measure of the 'goodness of fit' of the group of signals under inspection at any instant. Acceptable 'fits' greater than the 'fit' threshold may be produced when the group of signals is displaced one or more taps (radar periods) from the central position in the delay chain—however the best fit will be obtained when the group is centred.

To prevent all fit pulses becoming potential recognitions, the amplitude of the fit pulse is retained through the pulse width checking stage, and the pulse is passed

via a gate G to a delay unit—the delay of which is equal to one radar period. The output of this delay unit is compared with the input. If the input is greater than the output (i.e. this fit is better than that obtained one radar period earlier) then no 'best fit' pulse is generated. If the output of the delay unit is greater than its input then this input is inhibited by gate G and a 'best fit' pulse is generated.

8.3. Recognition Pulse Generation

The 'best fit' pulse suitably delayed is used to strobe the signal/background amplitude comparator and if the amplitude criterion is met during this strobing period then a 'recognition pulse' is generated.

An example of a typical beam shape obtained from an aircraft is shown in Fig. 12. The gated output of a logarithmic receiver has been photographed: the general background level due to receiver noise extends to some 15 dB above the mean logarithmic noise level denoted by the base line.

The downward vertical line following the signal group in Fig. 12 is a recognition pulse indicating that the group has been automatically detected.

8.4. System Performance

Briefly the following general results can be stated.

- (i) The automatic system has been shown to be no worse than alert human operators in initial detection performance.
- (ii) The automatic system will detect aircraft returns, after initial detection, with a probability

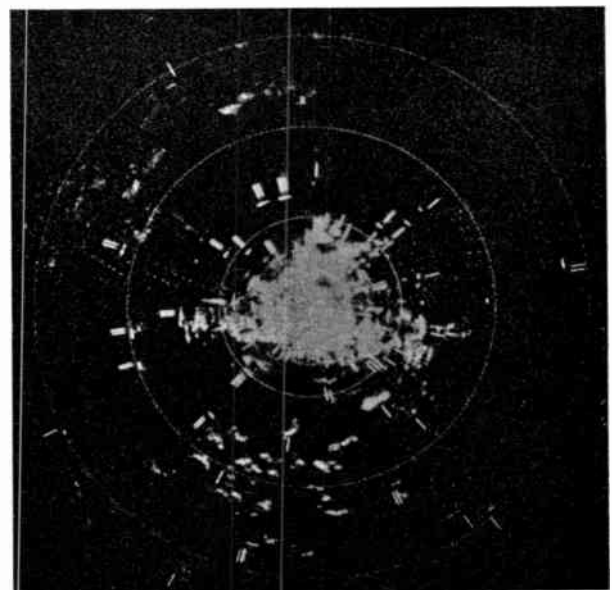


Fig. 13. Delay line auto-extraction system p.p.i. display of raw radar (1 revolution) with superimposed recognition pulses. Note the scarcity of detection in weather clutter areas.

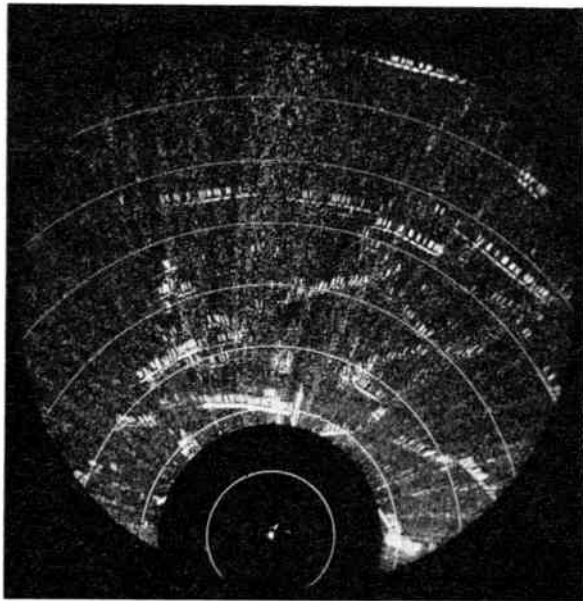


Fig. 14. Delay line auto-extraction system—off-centred radar picture with auto-detection markers (several aerial revolutions).

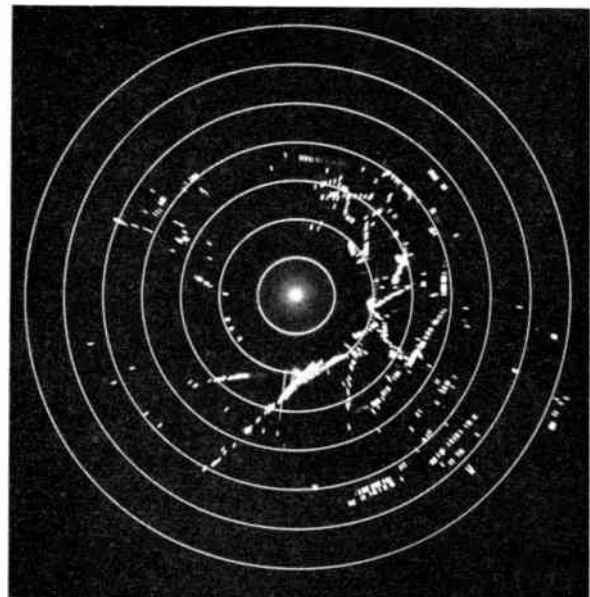


Fig. 15. Delay line auto-extraction system. Auto-detection markers only (several aerial revolutions).

sufficiently high to permit automatic track generation by an associated general purpose computer.

- (iii) The floating threshold provides a marked improvement in the rejection of unwanted clutter detections.

Auto-detection of raw radar returns is illustrated in Figs. 13, 14 and 15.

Figure 13 shows one radar revolution on a p.p.i. display. The radar picture shows heavy weather clutter conditions. The small radial markers are produced by the auto-detection circuits and in effect 'tick' every radar signal group that is automatically recognized. The scarcity of detection in the clutter blocks is illustrated.

Figure 14 shows an off-centred radar picture of several aerial revolutions with auto-detection pulses superimposed. This picture, and that of Fig. 15 which shows auto-detection pulses alone, illustrate the very real possibilities of automatic track compilation from auto-detected data.

9. Conclusion

The recognition pulses, as they are generated by the auto-detection equipment, gate range and bearing counts in digital form into an associated computer via a suitable buffer store which smooths the data

flow to the computer. The computer sorts the new data against data already received and compiles the necessary tracks. Thus the object of adequately matching the potential of the radar source to the capacity of an automatic data handling system is achieved.

It is believed that an auto-extraction/general-purpose digital computer system could cope with the normal radar environment adequately and also with a number of clutter and interference situations. However situations can be envisaged in which, with the present state of the art of auto-detection, man still has a useful part to play. The human being, with his highly developed sense of pattern recognition, is able to absorb and interpret the overall radar presented to him more readily than the machine. He is therefore suited to tasks such as inhibition of detections in selected areas, tracking selected targets through difficult environments and the rapid recognition of interference, etc.

The combination of an auto-extraction machine with limited human support is believed to provide an efficient and versatile radar data extraction system.

Manuscript first received by the Institution on 1st April 1963 (Paper No. 855/RNA21).

© The British Institution of Radio Engineers, 1963

(The discussion on this paper is printed on pages 327 and 328.)

DISCUSSION

Under the Chairmanship of Mr. A. St. Johnston

Mr. D. H. R. Archer: I believe that there is a logical error in Mr. Benjamin's calculation of the number of potential echo positions, since the validity of the assumption that each target will overlap a specified number of increments in bearing elevation and amplitude is dependent upon the truth of an implied assumption concerning the radar characteristics.

Be that as it may, the calculation itself seems to me to be of small value, since the discrepancy between the number of targets which a radar could handle and that which it is required to handle under any conceivable set of operational circumstances is so enormous. I should have preferred to hear an argument which took as its starting point a maximum practical situation. That, after all, is bad enough—but it may well be found that a practical solution exploiting the full *practical* capability of a radar would be less exacting than that required to exploit the full theoretical capability.

Mr. R. Benjamin (in reply): If we wish to detect all new targets promptly, and to extract all potential information concerning them, we must maintain an independent watch on each resolvable generalized 'space element' in which a new target might arise. In practice we may have a prior knowledge which restricts the space in which new targets may arise to only part of the surveillance volume. Also, at the expense of some degradation in detection performance, several intrinsically resolvable 'space elements' may be jointly observed for initial detection. At the further expense of some loss of time, the detailed information may then be abstracted afterwards, provided the target information is adequately correlated from one radar 'frame' to the next. However, few present-day surveillance radar data-processing systems get within striking distance of abstracting all the potentially available information on a target, once that target has been detected.

The section of the paper to which Mr. Archer refers, deals with the search for new targets. If we are concerned with tracking targets whose existence is already known, the relatively small number of those targets and the various 'redundancies' enumerated in the paper further ease the situation. However, even the exploitation of these factors does not permit the unaided human operator to realize anything approaching the flow of relevant and valuable information potentially available from the radar.

Mr. R. N. Lord: Man as a data processor is quite competent, his drawback is the effect of tiredness. The problem is well covered in the recent AGARDograph No. 60 "Man and Radar Displays" by C. H. Baker.

Mr. Benjamin (in reply): It is agreed that man, at his best, is extraordinarily good at pattern recognition and at making the best use of his limited rate of information acquisition, but that fatigue and inattention make it hard for him to maintain his optimum performance under conditions of prolonged routine duty.

Dr. D. E. N. Davies: On the subject of the potential data rates of both radar systems and humans as discussed

by Mr. Benjamin, there appear to be two approaches in trying to match the system and human together. The first, described by Mr. Benjamin is to filter the radar data and present only what is relevant to the human. An alternative approach might be to consider a radar system which had a far lower potential data rate. Clearly the resolving power of the radar could not be sacrificed but there is no real need in many radar applications to handle more than about 100 or 200 targets. This represents a large reduction of potential data rate. The properties of such a radar would then be similar to those of a radar data extraction system for use with narrow band radar data transmission. Furthermore if it were possible to realize such a low data rate system, the noise bandwidth could be theoretically reduced with a consequent potential for improvement in range performance.

Mr. Benjamin (in reply): The philosophy put forward by Dr. Davies is logical and attractive—but not often compatible with our present-day boundary conditions. The requirements of range performance, resolution of closely-spaced targets, anti-clutter performance, target-size discrimination and tracking accuracy normally prescribe the aerial gain (at the given frequency), and the same factors (except for range performance) may define the range-resolution. The need for the prompt detection of new targets, together with the requirement to establish and thereafter maintain the target velocity and track identity, continuity and accuracy may specify the data rate.

The segregation of detection from target analysis and tracking would permit a quasi-inertialess radar scan to match the radar operation, on those targets whose existence has been detected, to the configuration of these targets and to related operational criteria. However, this matching process itself would probably depend on automatic data extraction for its operation, and it is not directly applicable to the search for new targets.

Mr. G. D. Rhodes: The problem of radar capacity is important in calculating peak rates, but a completely bright screen conveys zero information in the same way as a completely blank screen. A 1 : 1 mark space ratio is the optimum and means that the maximum information is not equal to the total number of quanta but is divided by two for each dimension considered. Can the authors comment?

The amplitude, shape and pulse length have been mentioned as criteria for target detection. Can the authors indicate any numeric 'factor of merit' which can measure the overall effect of these variables?

Mr. Benjamin (in reply): Each independent space element returns an independent signal to the data processor. If all of them are empty—or full—this does indeed convey significant information.

If each space element were occupied by a discrete target, this would not result in a uniform energy density. Each target would produce a taper in amplitude over its resolution cell in range, bearing and elevation, reflecting the

transmitted pulse shape and the horizontal and vertical radiation patterns of the aerial (subject to various disturbing phenomena). Furthermore, the targets would not all be of the same amplitude. Finally, in the overlap zones, the instantaneous combined signal amplitude would vary, within each echo pulse, in accordance with the instantaneous r.f. phase relation between two partial returns differing in amplitude, fine range and Doppler velocity.

The limited dynamic range and resolution of a p.p.i. display may preclude the direct presentation of all these phenomena, even for a single-beam radar, but the information is nonetheless available at the output terminals of a radar receiver of adequate bandwidth. However, let us suppose that we are using a p.p.i., and that each echo produces a uniform arc of standard brilliance on this display; furthermore, let us assume that we require to resolve discrete echo centres (and not merely deduce that all resolution cells within a given area are occupied by targets). Even on this basis, adjacent echoes need be separated in bearing and range only by a 'black' gap corresponding to one p.p.i. spot diameter, not by one radar beamwidth or pulse length.

Finally, the paper discusses the problem in terms of the discrimination and resolution achieved by the radar. Surely the total discrimination (and similarly the total resolution) in orthogonal dimensions, is the product of the constituent discriminations (or resolutions) in each of these dimensions. Mr. Rhodes may wish to define these parameters in terms of the radar pulse length and modulation characteristics and its beam shape and angular pulse spacing. But this has no direct bearing on the validity of the argument presented in the paper.

Mr. J. C. Plowman (in reply): A 'figure of merit' as such for the detection probability of weak signals in noise is not available. Obviously the signal to noise ratio for, say, 50% probability of detection with any detection system—human or otherwise—depends on the parameters of the radar involved, and auto-detection systems have been shown to be no worse than human operators in this respect. Beam shape and pulse length are useful to reduce clutter and improve accuracy, they do not contribute positively to detection against a background of Gaussian noise.

Mr. P. R. Joanes: One must be extremely grateful to Mr. Plowman for his admirable discourse on the important aspect of automatic radar data extraction and his choice of subject matter in the time available. Nevertheless, I am sure he would agree with me that since the practicability of the system depends on the technological considerations governed by the physics of the delay lines, there is some interest in pointing out to the Symposium, some of the difficulties that have been met. In a system of cascaded delay lines, considerations apply which are similar to that of a communication system. Among these difficulties may be listed, the requirement of frequency multiplexing in a somewhat arbitrary transmission medium, the preservation of signal to noise ratio, the possibility of cross talk at v.h.f. across a complex system within a confined space and the need for very strict linearity and stability conditions for an analogue device working in conjunction with a

digital computer. By no means the least consideration has been the admirable and unique compromises in delay line design by Mr. C. F. Brockelsby and his team.

Mr. W. R. Daniels: In view of the general trend in electronics for smaller equipment and less power consumption, how do storage systems by means of delay lines and storage tubes compare?

Mr. Plowman (in reply): Size and power consumption have not been major items for consideration in the investigations described. The choice of auto-detection system is more affected by the parameters of the radar involved than by size and power consumption factors—both of which are normally insignificant compared with those of the radar installation. Operationally, storage tubes have poorer life, resolution and stability than delay lines, but they need less specialized development, do not require a stable pulse repetition rate, and are adaptable to a wide range of beamwidth, scanning speed and pulse repetition rate.

Mr. G. J. Jones: In the case of a minimum detectable fluctuating target, what advantage does Mr. Plowman claim in using the beam shape as one of the criteria for detection, compared with the system of using only amplitude and beamwidth as the criteria?

Mr. Plowman (in reply): It is considered that the beam shape criteria enables better bearing accuracy to be achieved than is possible with a 'beamwidth' criteria. The beam shape criteria also does provide a marked discrimination against large clutter signals and further it has been found that shape variations of large signal groups from aircraft, when examined on a logarithmic scale, are not as significant as might be expected, and these shape variations result in an insignificant loss in probability of detection when visual detection is 100%.

Mr. R. N. Lord: Would Mr. Plowman indicate the effect of requiring several successive pulses for recognition on the probability of detection of the target?

Mr. Plowman (in reply): Acceptance of a single strong pulse as a target would leave the system excessively vulnerable to impulsive interference.

Dr. D. E. N. Davies: The system described by Mr. Plowman correlates the envelope of the received pulses with the two-way directional characteristic of the aerial. I should like to ask Mr. Plowman whether there is any advantage to be gained in the form of improved angular accuracy, from the use of a split-beam type of directional characteristic as used in some direction-finding systems.

Mr. Plowman (in reply): Regarding Dr. Davies' question, I am afraid I am not acquainted with the direction finding systems to which Dr. Davies refers—however if the system provides a split beam which rotates *past* a target position and hence produces an envelope of returns which has a 'shape' determined by the aerial radiation pattern then I do not believe that improved bearing accuracy will be achieved.

If, however, the split beam technique is employed in a tracking radar then beam shape auto-detection techniques cannot be used, since the beam does not sweep through the target position.

Propagation of Sound in Shallow Water

By

D. E. WESTON †

Presented at the Symposium on "Sonar Systems" in Birmingham on 9th–11th July 1962

Summary: The way the medium controls sound reception in shallow water is discussed generally, and illustrated by some North Sea experiments with explosion sources. The ray and mode theories are introduced, and it is shown where it is convenient to change from one to the other. Transmission loss is the most important parameter, and is least around 200 c/s. Time dispersion is mainly due to the dispersion in the vertical arrival angle. Large fluctuations are observed when using a continuous wave source. The propagation is dominated by the characteristics of the two boundaries.

1. Introduction

This paper describes the way the shallow-water medium controls the reception of sound; i.e. such characteristics as transmission time, transmission loss, arrival angles, the dispersion in the arrival time of a pulse, and the fluctuations in the level of a pure tone signal. It is convenient to define water as 'shallow' if the depth is less than 100 fathoms and this paper is intended to be a practical account which points out some of the things of importance, but, for lack of time, does not go into much detail nor very deeply into theory. Much of it is not new. It is essential to include examples, and most of these are taken from some early North Sea experiments.

2. North Sea Experiments

The experimental results will be stated as relevant through the paper, but it is worth introducing the experiments briefly. An area of exceptionally constant water depth (about 40 fathoms) was chosen on the Great Fisher Bank in the Central North Sea, running north from 56° 05' N, 03° 15' W. It was hoped that the area would show a simple behaviour acoustically, which could be used to check various theories of transmission, but the hopes for simplicity were not fully realized. It was visited three times in 1954, in order to cover a range of thermal conditions. R.R.S. *Discovery II* acted as the receiving ship, and the acoustic sources were mostly small underwater explosions fired by one of H.M. ships. Sound reception was investigated as a function of range and depth, concentrating on the eight octaves in the frequency range 25 c/s to 6.4 kc/s. These frequency limits are mainly set by the source, the explosion energy is relatively small at higher frequencies and the source behaviour is complicated at lower frequencies.¹ Transmission varies considerably as one goes through this frequency range, so that no one octave is truly

† Admiralty Research Laboratory, Teddington, Middlesex.

representative—however the examples here have been drawn from one of the middle octaves (400–800 c/s referred to as 560 c/s). Many auxiliary observations were made. It is hoped soon to publish a full report on these propagation experiments and on the North Sea bottom characteristics.

3. Simple Theories

Sections 3.1 and 3.2 outline the well-known theories which are compared in Section 3.3.

3.1. Ray Theories

The best-known solution of the wave equation is in terms of rays. In an ideal ocean; which is unbounded, homogeneous and loss-free, sound spreads out spherically. The source and the receiver are connected by a single straight-line ray. The next stage of complication is that there can be multiple rays, due to reflections from the surface and the bottom of the sea. If the sea is inhomogeneous (e.g. layered) the rays will be curved, and in fact the concept of a ray is now only an approximation—though often a good one. In general there are losses, both in transmission through the medium and on reflection from the boundaries. The bottom interface may sometimes be treated as the plane boundary between two ideal fluids, when the Rayleigh reflection laws will be obeyed. Provided the sound velocity in the bottom material is greater than that in the water there is no loss at small grazing angles, but beyond a critical angle the losses rise rapidly. The energy incident on the bottom at angles below the critical angle is trapped, so that there is a channelling or guiding of this energy to give a cylindrical spreading law.

3.2. Mode Theories

At long ranges there are a large number of ray arrivals, and in addition the phase relations between the arrivals become important. A more convenient form of the solution of the wave equation is in terms

of the normal modes of propagation. It is necessary to remember that there is a phase change of π at the free surface, with a lesser but variable change at the bottom. The classic treatment of this problem is due to Pekeris.² The approximate treatment described here is possible because, for angles up to the critical angle, the bottom interface may be replaced by an equivalent free surface lying just below the real bottom.³ The distance below the bottom is almost independent of angle, and typically equal to about three-quarters of a wave-length. However the distance is not independent of frequency, which must be remembered in any group velocity calculations. (This is most important for the ground waves, which will not be discussed.) Consider therefore the wave equation for propagation between two parallel free surfaces, taking the two-dimensional case but remembering there is really an additional cylindrical spreading term. The solution may be obtained as the sum of a number of modes, and there is a limited number of discrete modes which have real phase constants. Mode amplitude varies sinusoidally with depth, the n th mode has n pressure maxima. A given mode may also be represented by the superposition of two plane waves of grazing angle ϕ , one upgoing and one downgoing. On reflection one of these waves is converted into the other. The phase velocity of a mode is greater than the free space velocity c ; it is given by the velocity of the point of intersection of one of the equivalent plane waves with one of the boundaries, i.e. $c \sec \phi$. The angle ϕ is related to wavelength λ and enhanced water depth H by

$$\sin \phi = \frac{n\lambda}{2H} \quad \dots\dots(1)$$

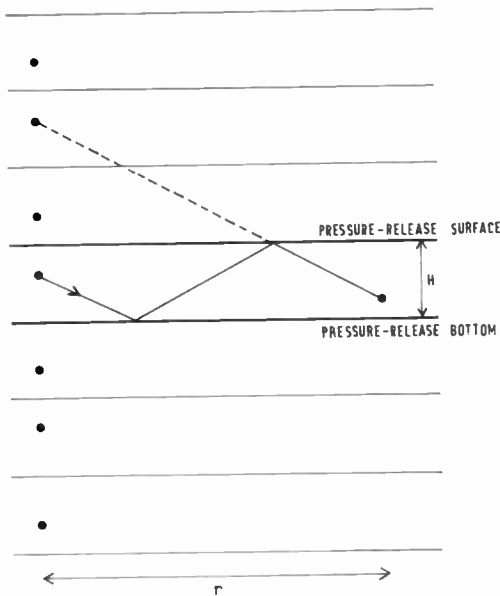


Fig. 1. The image concept.

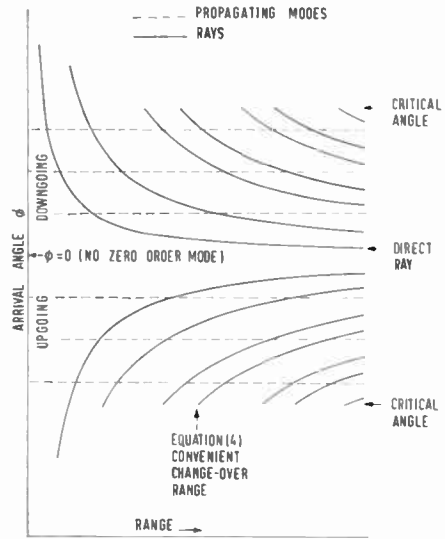


Fig. 2. Arrival angles.

The above mode theory is in its simplest form, but many workers have shown that it is possible to extend it to cases where the sound velocity depends on depth. It is also possible to extend either ray or mode theory to the case of varying water depth.⁴

3.3. Choice of Ray or Mode Concepts

For a given shallow water situation there is the choice of using either the ray or the mode concept. In the literature the advice given is only qualitative—to use rays for deep water, high-frequencies and short ranges; and to use modes for shallow water, low frequencies and long ranges. It is desirable to make this quantitative, and a special case will be treated initially, having iso-velocity water between two plane parallel free surfaces. Figure 1 shows a third way of looking at this problem, in terms of the images due to reflections at the surface and bottom. The polarities of the images alternate, due to the phase changes on reflection. The line of images acts rather like a simple line source, insofar as both tend to produce cylindrical spreading. It may be seen that the average difference in arrival angle between neighbouring low-angle rays is

$$\Delta\phi \simeq \frac{H}{r} \quad \dots\dots(2)$$

where r is range.

The line of images also behaves like a line array of projectors, or a diffraction grating of spacing $2H$, in that there are certain preferred directions in which all the images add coherently. These directions are given by eqn. (1), and are in fact the directions of the equivalent plane waves which constitute the modes. The average difference in angle between neighbouring low-order modes is

$$\Delta\phi \approx \lambda/2H \quad \dots\dots(3)$$

The angles for the rays and modes are drawn in Fig. 2, note that in general there is no simple relation between the two sets of angles. This is because the ray and mode concepts are quite different in nature: rays imply a source at some finite distance and curved wavefronts, whereas modes are essentially plane wave concepts valid asymptotically at very long ranges.⁵ If one adds up the ray or image contributions an exact solution is obtained, though for short ranges the effect of the higher-order images is very small and the practical calculation is manageable. At long ranges the necessary number of images is very large, and since it is strictly necessary to take account of phase the calculation becomes unmanageable. Similarly an exact solution can be obtained by adding up all the modes (including in the general case Pekeris' branch-line integral), but at long ranges in practice only the propagating modes need to be considered. Close in the mode solution contains many more significant terms and becomes unwieldy. Thus in this special case both the ray and mode solutions can be exact, both can in fact be transformed into the other, and the choice between them is only a matter of convenience or of minimizing computation.

Equations (2) and (3) show that the average angular spacing of low-order arrivals is the same for ray and mode concepts at a range given by

$$r = \frac{2H^2}{\lambda} \quad \dots\dots(4)$$

As already discussed each mode has two angles and two plane waves associated with it, but there is only one arrival. The range for approximate equality in the number of low-angle arrivals is therefore

$$r = \frac{H^2}{\lambda} \quad \dots\dots(5)$$

The lower-angle arrivals are the most important ones, in the practical case the effective angular limit may occur at the critical angle for both rays and modes. Thus below the range limit of eqn. (4) or (5) there are less effective ray terms than mode terms, and *vice versa* above the limit. It is convenient but not obligatory to change over from the idea of rays to that of modes at one of these limits. Since it is only a matter of convenience it is permissible to use either the eqn. (4) range based on the angle criterion, or the eqn. (5) range based on numbers of arrivals. The choice depends on the purpose, and although the latter is preferred there could be a bias towards eqn. (4) since ray calculations are easier than mode calculations.

As the source and receiver depths are varied there may be occasions when certain ray arrivals come in at the same time, and also occasions when certain modes are not excited (there is no permitted zero-order mode at all with two free surfaces). There is also the coinci-

dence that with source and receiver at half-depth the ray and mode angles and arrival times are identical at the range of eqn. (4), provided the low-angle approximation is used. In the practical case too there is a real sea-bed, real sea-surface and a real water structure to contend with, and there is a discussion on observed time dispersions in Section 5. However, the eqn. (5) change-over range is only one of convenience, so that its use in most circumstances is permissible.

If either the ray or mode approach are used with approximations there may be other limits where it is not only convenient but obligatory to change over to the alternative approach. For example a ray method with incoherent addition of arrivals cannot be used when all the effective ray angles are less than the angle for the first mode.

It may be seen that eqn. (5) is almost identical with that giving the limit of the Fresnel zone surrounding a linear array of length *H*. It may also be expressed in other ways, e.g. as the discriminant $H^2/\lambda r$ or as the critical water depth

$$H = (\lambda r)^{\frac{1}{2}} \quad \dots\dots(6)$$

It could be said that eqn. (6) provides a better definition of acoustically shallow water than the 100 fathom limit quoted above. However 'shallow-water propagation' is not identical to 'mode propagation' and one sometimes gets the typical shallow-water-type propagation even for depths greater than the eqn. (6) limit, though it is always necessary that there should

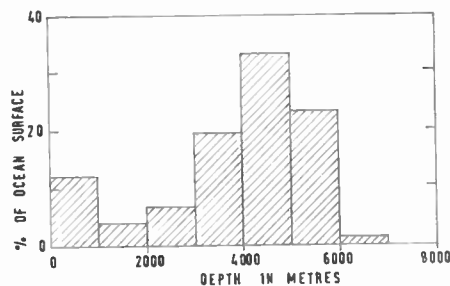


Fig. 3. Histogram showing distribution of ocean depths.

be effective paths involving very many bottom bounces. Acoustically the usual distinction between deep and shallow water is rather unsatisfactory. The real justification for the distinction comes from the bimodal distribution of depths in the ocean rather than from the physics of the problem (see Fig. 3 based on ref. 5). Intermediate depths are encountered on the Continental Slopes and Continental Rises, but form only a small part of the total oceanic area.

Take as an example 560 c/s sound in the North Sea area. The wavelength is about 9 ft, and under iso-velocity conditions the angle ϕ for the first mode is just over 1 deg. The eqn. (5) range of convenience for

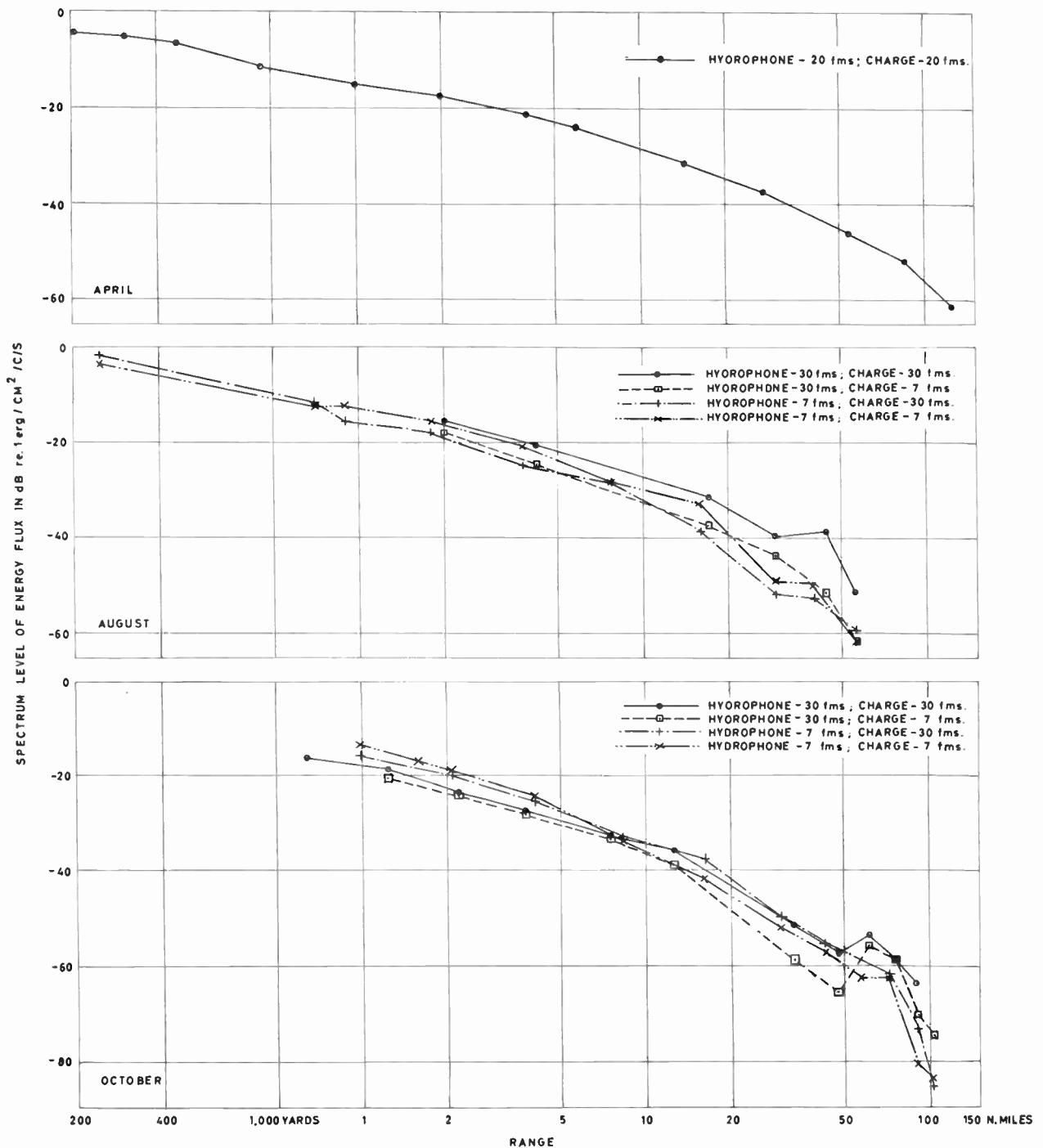


Fig. 4. 560 c/s transmission curves for the North Sea.

the ray-mode change-over is about one nautical mile. From refraction shooting the bottom velocity is about 1.17 times the water velocity, corresponding to a bottom critical angle of about 30 deg. Presuming a smooth sea surface there should be 25 possible propagating modes.

4. Transmission Loss

To a first approximation the levels at long ranges may be predicted by assuming spherical spreading out to some range and then cylindrical spreading. The change-over range is given by $H/2\phi_{crit}$ where ϕ_{crit} is

the effective critical angle.⁴ In addition there is the bulk absorption in the medium, and the attenuation due to second-order losses at an absorbing bottom, due to a coupling into shear waves at rigid rocky bottoms, and due to scattering at both boundaries. The magnitude of the losses is very important, and there is some discussion in the literature (see especially refs. 7 and 8 for the losses at an absorbing bottom). Note that the higher modes will usually lose energy faster than the others, both because they have a greater grazing angle at the bottom and because their equivalent plane waves hit the bottom more often. If there are many modes present it may be shown that a mode attenuation proportional to ϕ^m should lead to a transmission law of intensity proportional to $r^{-(1+1/m)}$.

The proof for this relation may be briefly indicated by considering transmission from r_1 to r_2 in a medium of constant depth. The cylindrical spreading or r^{-1} law will multiply the intensity by the factor r_1/r_2 . It will also be multiplied by ϕ_2/ϕ_1 where ϕ is the range of angles effective in carrying energy. At r_1 the limit of effective angles ϕ_1 will be determined by the angle at which there is appreciable attenuation, and approximately this same attenuation should determine the limit ϕ_2 at r_2 :

$$r_1 \phi_1^m = r_2 \phi_2^m \quad \dots\dots(7)$$

The total attenuation may be written

$$\frac{r_1 \cdot \phi_2}{r_2 \cdot \phi_1} = \left(\frac{r_1}{r_2}\right)^{(1+1/m)} \quad \dots\dots(8)$$

One example of the law is for $m = 2$ when intensity is proportional $r^{-3/2}$. The author considers however that there is no experimental evidence for such laws holding over any appreciable range interval. The only power laws for which there is much evidence are the cylindrical spreading law, spherical spreading law, and perhaps the dipole r^{-4} law. The spherical law, the ubiquitous 6 dB per distance doubled, turns up for a variety of reasons in many places where it is not expected—and where it has no right to be!

Note that shallow-water losses are greater in summer than in winter. This is because summer heating causes conditions of downward refraction, leading to a larger number of bottom bounces at steeper angles. These seasonal effects have in fact been known since the first World War, from observations on underwater bells.⁹

In Fig. 4 some 560 c/s measurements are presented, showing the dependence on range at three seasons of the year. The density of energy flux for an octave band is measured and reduced to a spectrum level; this quantity for a charge should be comparable to intensity spectrum level for a continuous source. In Fig. 5 the results have been converted to transmission loss

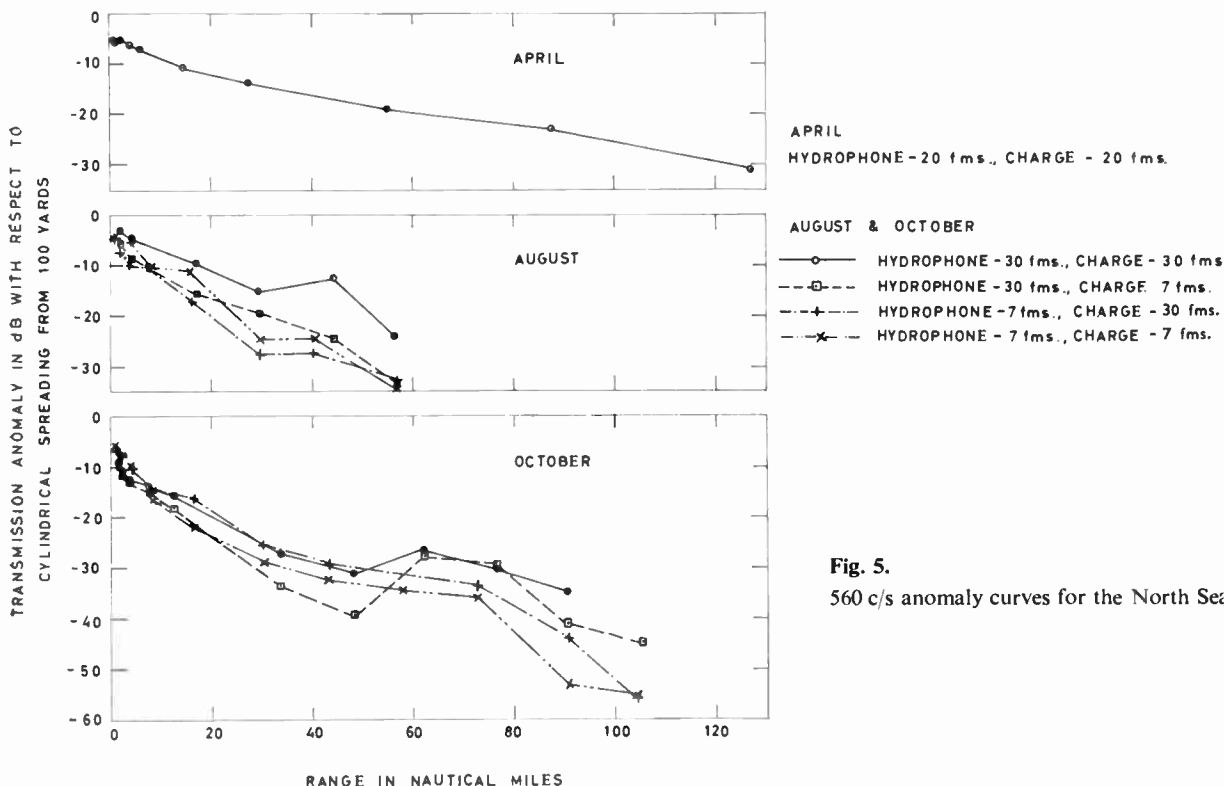


Fig. 5.
560 c/s anomaly curves for the North Sea.

by assuming a 560 c/s source level of +3.2 dB rel. 1 erg/cm²/c/s at 100 yd, and then corrected for cylindrical spreading. Note first the seasonal differences, to be discussed in section 8. Next look at the shapes, which are not too far off straight lines. If one forces a straight line fit for ranges beyond about 5 miles, the attenuations deduced vary from only 0.18 dB/mile in April with isothermal conditions to 0.48 dB/mile in August with strong layering. For the reasons to be given in Section 7 it is dangerous to read too much into the shape of transmission curves, but the Fig. 5 curves are surprisingly linear. There is only a slight tendency for them to curve upwards, corresponding to a reduced attenuation at the longer ranges where the higher modes have been stripped off. With other evidence this had led to the idea that there is a continual interchange of energy amongst the modes due to scattering. At long ranges this can lead to a dynamic equilibrium in the distribution of energy amongst the modes, and this system will have an attenuation coefficient which does not change with range. It is hoped to publish the evidence for the redistribution theory soon.

5. Time Dispersion

The spread of arrival angles shown in Fig. 2 suggests the type of dispersion to be expected. The relative delay in arrival is $r(1 - \cos \phi)/c$ for rays, and also holds for modes in the special case having two free surfaces. At close ranges the ray picture should hold, suggesting a series of discrete arrivals with little fre-

quency dispersion within each arrival. For the longer ranges the mode picture also suggests discrete arrivals, but since the group velocity depends on frequency there is appreciable dispersion within each arrival—and they will usually overlap. The frequency-time curves for up to four mode arrivals from a single shot have been displayed on the Sonagraph analyser by Hersey *et al* and by Ewing *et al.*¹⁰ There is also likely to be scattered energy coming in between the mode arrivals. At first sight Fig. 2 suggests that dispersion will be proportional to range, but in practice there is a much slower increase with range because of the selective attenuation of the higher modes. Dispersion actually depends on range, number of modes received, and effective bandwidth for each mode.

Dispersion may be most vividly illustrated by playing back recordings of shot signals, such as those in the North Sea. There is a regular change with range, starting at zero range with a shot that really illustrates reverberations. In the isothermal conditions of April the sounds are not particularly unusual. However in layered conditions, such as those shown in Fig. 6 for August, the signal may be drawn out into a long and weird moan quite unlike a shot. This is because one range of modes and frequencies travels with a group velocity close to the free-space velocity beneath the thermocline, and another range travels with a velocity close to that above the thermocline.

Obviously time dispersion limits the range resolution in an echo-ranging system. It is also generally

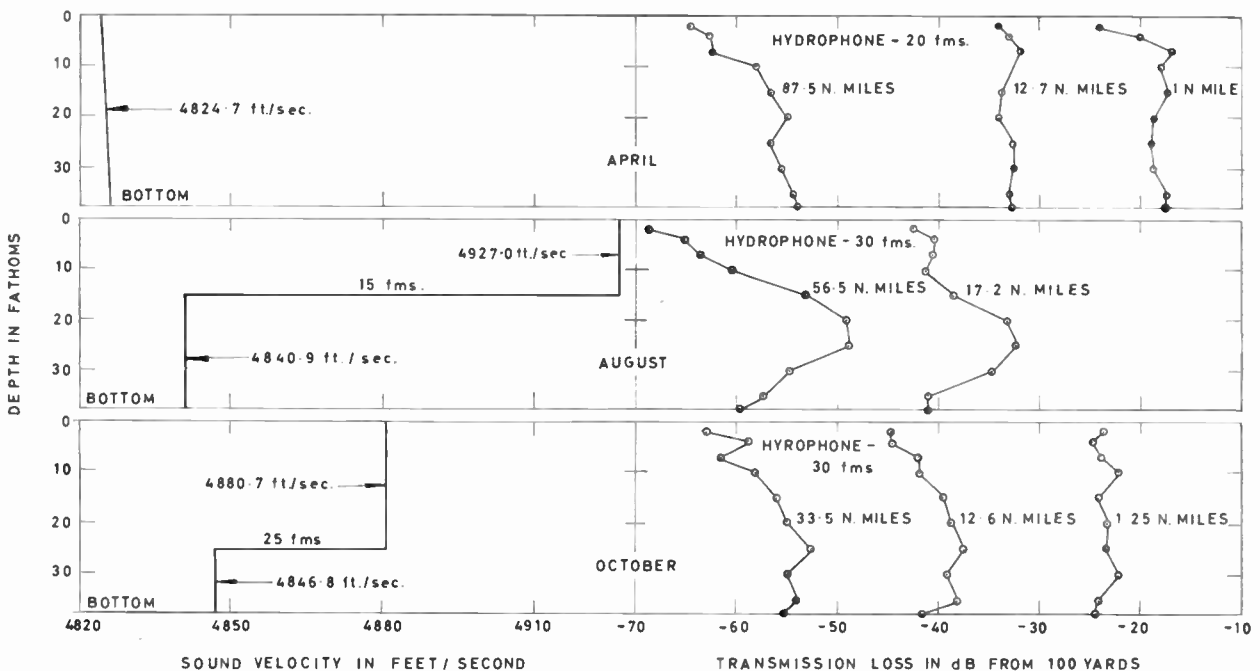


Fig. 6. Idealized sound velocity profiles and 560 c/s depth-dependence curves for the North Sea.

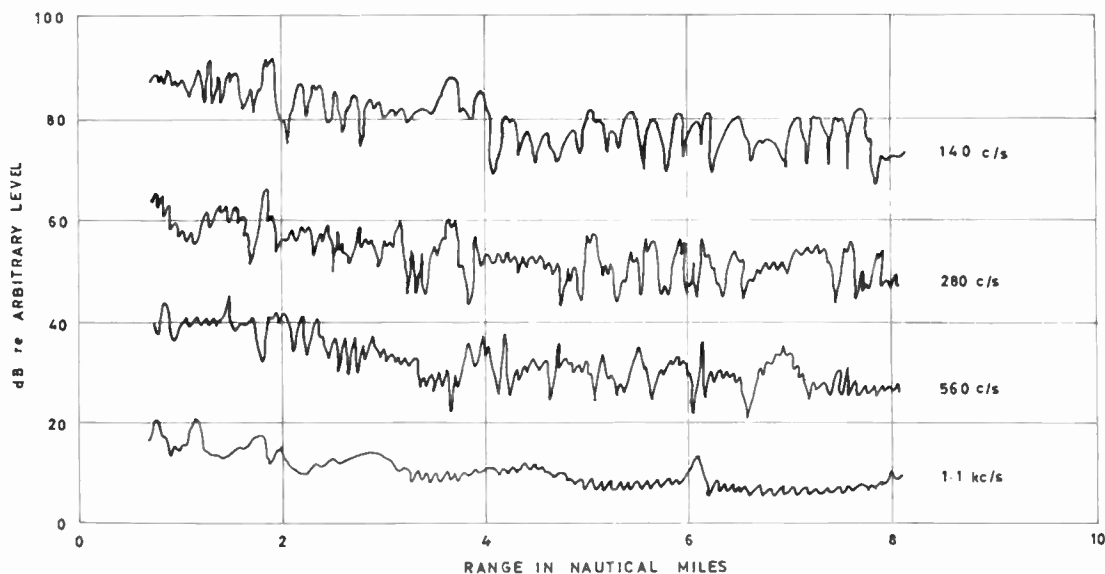


Fig. 7. Hammer-box transmission curves for the North Sea in April.

true that vertical dispersion in arrival angle will limit the horizontal angular resolution. For shallow water it is best to think of phase velocity, which increases with mode number. Thus an electrically-steered linear array looking at an off-axis source will receive a number of maxima as time delay is varied, each corresponding to a mode (compare ref. 11). In a medium of variable depth there are related dispersions in both vertical and horizontal arrival angles.¹²

It is considered that the study of time dispersions has been comparatively neglected, and should prove a useful attack point in attempting to understand propagation mechanisms.

6. Fluctuations

If a pure tone is transmitted the signal received at a distance will nevertheless show large fluctuations in both amplitude and phase, which may be described alternatively as a frequency spread of the original monochromatic line. To produce fluctuations it is necessary that there should be either scattering or

multiple path interference; plus movement of source or receiver or medium; with either regular or random movement. Some good shallow-water fluctuation measurements are described by Mackenzie.¹³

Fluctuations are most marked for pure tone sources, but may still be large for broad-band measurements on sources containing lines, or even for sources not containing lines. Figure 7 shows the sizeable fluctuations for one typical transmission experiment with a hammer-box source in the North Sea, even with a little smoothing and with an octave bandwidth. Despite fluctuations the shapes of the transmission curves from the hammer-box intensity measurements and the explosion energy measurements agreed reasonably well. A sample comparison adjusted vertically for best fit is shown in Fig. 8.

7. Influence of the Sea Surface

It is important to realize that in shallow water the propagation of sound is dominated by the characteristics of the two boundaries. Of these two the bottom

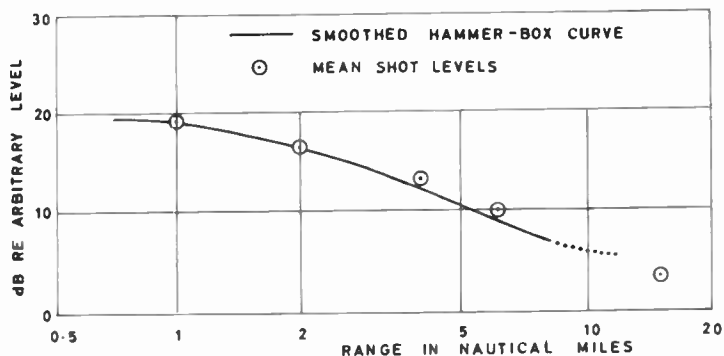


Fig. 8. Typical comparison of shot and hammer-box runs for the North Sea in April. Frequency 560 c/s.

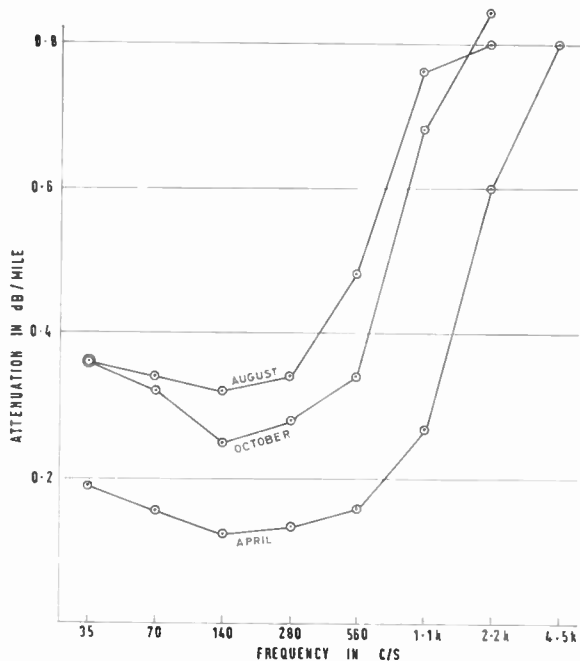


Fig. 9. Mean attenuation coefficients for the North Sea.

is the more important because it can vary more, but let us take surface effects first. The most obvious surface effect is due to the phase change on reflection. If either source or receiver is near the surface there is the surface dipole or Lloyd's mirror phenomenon. The loss near the surface may be seen in Fig. 6, the effect is more marked at long ranges where the average grazing angles are smaller.

Another important effect is the change-over from specular to diffused reflection as the grazing angle and frequency increase.¹⁴ In the North Sea experiments the transmission loss depended on the sea conditions—significantly on the wave height and probably on the wave direction as well. This appeared to hold for the whole frequency range 25 c/s to 6.4 kc/s. For example in Figs. 4 and 5 the dip in the October results near 50 miles could be due to the fact that measurements at this range were taken in rough conditions. This means that neither the shape of the transmission curve, nor the observed attenuations, are necessarily typical of an area if they are deduced from a set of measurements lasting a few days.

8. Influence of the Sea Bottom

It is found that shallow water areas tend to split up into two types. They have similar transmission losses at the higher frequencies but show large differences at the lower frequencies—e.g. 50 dB difference at 5 miles for 35 c/s. Observations have shown that areas with a high loss at low frequencies have at the bottom either exposed rock, or rock covered by only a few feet of

sediment. There is a large reflection loss due to shear wave coupling. These areas also tend to show particularly large seasonal effects. The other type of area has bottoms of sand, mud, shell, etc.—and the North Sea area with a sand bottom is typical. Figure 9 shows for three seasons the attenuation (measured beyond 5 miles) as a function of frequency, it is least for the 140 c/s octave. The best frequency for high-loss areas would typically be an octave or so higher.

It must not be assumed that the North Sea bottom behaves simply. The August and October results in Fig. 6 show a 'bottom dipole' effect which is not present in April. With some other results this makes it probable that there are large quantities of gas at the bottom in the summer months.

The detailed shapes of the depth-dependence curves in Fig. 6 have not been discussed, but it may be remarked that they correlate with the sound velocity profiles shown in the same figure, and can be used for quantitative deductions on the number of modes present and on the character of the bottom.

9. Acknowledgments

The North Sea experiments from which some examples have been taken were made possible only by the helpful co-operation of the Captains and Ships' Companies of several ships, and the efforts of many colleagues at the Admiralty Research Laboratory.

10. References

1. D. E. Weston, "Underwater explosions as acoustic sources" *Proc. Phys. Soc. London*, 76, pp. 233-49, August 1960.
2. C. L. Pekeris, "Theory of Propagation of Explosive Sound in the Ocean", (Geological Society of America Memoir No. 27, October 1948).
3. D. E. Weston, "A moiré fringe analog of sound propagation in shallow water", *J. Acoust. Soc. Amer.*, 32, pp. 647-54, June 1960.
4. D. E. Weston, "Guided propagation in a slowly varying medium", *Proc. Phys. Soc. London*, 73, pp. 365-84, March 1959.
5. H. U. Sverdrup, M. W. Johnson and R. H. Fleming, "The Oceans", (Prentice-Hall, New York, 1942).
6. I. Tolstoy, "Modes, rays and travel times", *J. Geophys. Res.*, 64, pp. 815-21, July 1959.
7. K. V. Mackenzie, "Reflection of sound from coastal bottoms", *J. Acoust. Soc. Amer.*, 32, pp. 221-31, February 1960.
8. K. V. Mackenzie, "Long-range shallow-water transmission", *J. Acoust. Soc. Amer.*, 33, pp. 1505-14, November 1961.
9. F. Aigner, "Unterwasserschalltechnik", (Berlin, 1922).
10. M. Ewing, S. Mueller, M. Landisman and Y. Sâto, "Transient analysis of earthquake and explosion arrivals", *Geofis. Pur. Appl.*, 44, pp. 83-118, 1959.

11. C. S. Clay, "Array steering in a layered waveguide", *J. Acoust. Soc. Amer.*, **33**, pp. 865-70, July 1961.
12. D. E. Weston, "Horizontal refraction in a three-dimensional medium of variable stratification", *Proc. Phys. Soc. London*, **78**, pp. 46-52, July 1961.
13. K. V. Mackenzie, "Long-range shallow-water signal-level fluctuations and frequency spreading", *J. Acoust. Soc. Amer.*, **34**, pp. 67-75, January 1962.
14. H. W. Marsh, M. Schulkin and S. G. Kneale, "Scattering of underwater sound by the sea surface", *J. Acoust. Soc. Amer.*, **33**, pp. 334-40, March 1961.

Manuscript first received by the Institution on 8th June 1962 and in final form on 22nd January 1963 (Paper No. 856/SS1).

© The British Institution of Radio Engineers, 1963

POINTS FROM THE DISCUSSION

Mr. M. Schulkin†: Dr. H. W. Marsh and I have sent a letter to the editor of the *Journal of the Acoustical Society of America* on "Shallow-Water Transmission" in which we have analysed a very large quantity of shallow water data and have presented expressions for the computation of transmission loss.‡ In general, we consider three zones: (1) the near inverse square spreading zone, (2) the far cylindrical spreading zone, (3) a zone between these two with $r^{-3/2}$ spreading. In line with the present result that $r = H^2/\lambda$ separates the region between facile ray treatment and facile mode treatment, we find empirically that $r = 8H$ furnishes a suitable division, i.e. $8 = H/\lambda$.

The author (in reply): In my paper I did not attempt to present equations suitable for computation, but I see no reason why your empirical expressions, with division into zones, etc., should not be satisfactory. I note that the $r^{-3/2}$ law in the intermediate zone is different in character from the one discussed in my Section 4.

† Westinghouse Electric Corp., Baltimore, U.S.A.
‡ *J. Acoust. Soc. Amer.*, **34**, pp. 363-4, June 1962.

Mr. M. J. Daintith§: Do not the curves of Figs. 5 and 8 fit an $r^{-3/2}$ law very well, in spite of the statement that this does not occur in nature?

The author (in reply): I consider there is great difficulty in establishing (or refuting) any given transmission law. First, one can find a portion of almost any experimental curve which will follow the given law, and secondly one can find isolated cases where there is good agreement over an appreciable range. For a proper test one needs a large quantity of data covering a wide range interval, and perhaps a statistical approach. Only sample data are presented in the paper, but consider nevertheless Fig. 4—which is the best figure to examine since it shows all the data in a form in which the $r^{-3/2}$ law will appear as a straight line. Middle values of range may be chosen where the slope does follow the $r^{-3/2}$ law, but I strongly contend that taking all the curves there is no evidence for linearity in this region. I can also repeat that I have never seen acceptable evidence for a real $r^{-3/2}$ law elsewhere, though I would not have been surprised to have found some.

§ Admiralty Underwater Weapons Establishment.

Test and Display Equipment for Multi-aperture Shift Registers

By

P. BROOMER (Graduate)[†]

Summary: The mode of operation of multi-aperture ferrite core devices is explained briefly and applications as shift registers discussed. An equipment for displaying the characteristics of the registers is described.

1. Introduction

The multi-aperture device employs specially-shaped magnetic cores utilizing square hysteresis loop ferrite materials in which a number of minor apertures are spaced about a major aperture. The shape of a typical core and the flux path indicated by an arrow, when the core is in the '0' condition is shown in Fig. 1(a). In all logic systems it is essential to have discrimination between the '0' and '1' conditions; an example of this is a transistor which can be in the 'on' (conducting) or in the 'off' state. With cores which can have more than one flux path as shown in Fig. 1(b), there are two flux paths in opposite directions around the core. Comparing this with the flux in Fig. 1(a) we have a basic difference in the state of the cores. The core which has two flux paths in opposite directions is said to be in the '1' condition. This then gives us the logic elements we require—one core in the '0' and one in the '1' state.

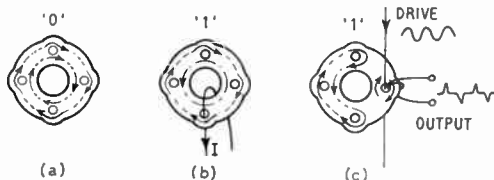


Fig. 1. Digital setting and non-destructive read-out.

To obtain information from multi-aperture cores without destroying the information already held in the core a minor aperture must be pulsed with an a.c. signal. If the core is set in the '1' condition the a.c. signal will change the flux about the minor aperture in sympathy with the drive circuit, hence an output will be detected by the output winding threaded through the same minor aperture (Fig. 1(c)). In the '0' condition of the core all the flux is saturated in one direction and thus the drive current will tend to saturate the core even more and no output will be detected as there will be no flux change around the

minor aperture of the core. When obtaining information from the core there is no change in flux within the core, thus information as to the state of the cores can be obtained without the re-write circuits of single aperture cores. The amount of power available from the core depends on the rate of change of flux around the minor aperture; the cores are capable of giving 100 mW which is sufficient to light a small incandescent lamp. This simplification of detecting the state of the core by means of visual display gives condition whereby if the lamp fails when the core is in the '1' condition the lamp will not light, thus forming a 'fail-safe' device. Trigger pulses may also be obtained from the cores by replacing the a.c. drive with a positive and negative square loop pulse which will give the state of the core at any time.

When information is required to be transmitted from one core to an adjacent core, as in a shift register, coupling loops through minor apertures of the cores are required. With reference to Fig. 2(a), two cores are shown coupled by two minor apertures. The first core is set with a '1', no flux being transmitted to the adjacent core as the flux in the transmitting minor aperture is not changed; thus no current flows in the coupling loop. A prime winding is included in the cores which reverses the flux locally around the

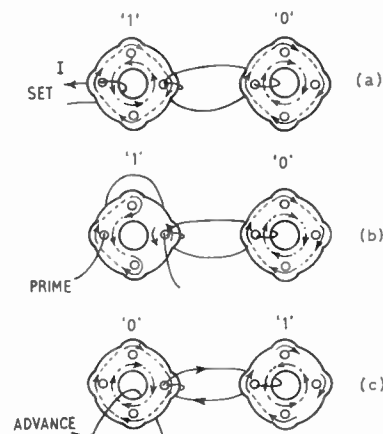


Fig. 2. Diode-less transmission.

[†] Aircraft-Marine Products (Great Britain) Ltd., 87-89 Saffron Hill, London, E.C.1.

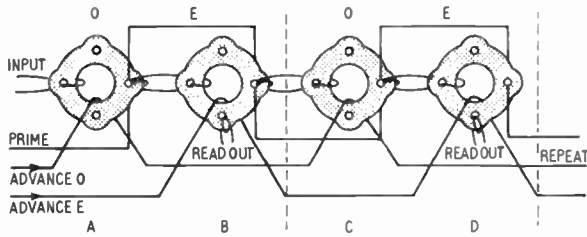


Fig. 3. Schematic shift register circuit.

transmitting minor aperture of the first core. This prime current is held below the minor/major aperture threshold so no signal, which will influence the setting of the adjacent core, is induced in the coupling loop. A clearing or advance pulse is then applied to the first core through the major aperture as shown in Fig. 2(b). Since the current is large and is so shaped that the flux changes about the minor transmitting aperture are large, the coupling loop will have a current of sufficient magnitude to 'set' the second core (Fig. 2(c)). The information set in the first core has now been transmitted to the second, and by coupling several cores together, using two cores per bit, a shift register or other logic device may be constructed.

The wiring of part of a shift register is shown diagrammatically in Fig. 3, where two cores are wired with an advance winding O to E and two cores with advance E to O, the prime winding being common to all four cores. Two cores per bit are required as it is impossible to load the device with coincident '1's without using two cores. When transmission is taking place between cores A and B, no input into the minor aperture of core A would have any effect as the cores have only two states, thus, it is essential to have two cores per bit.

2. Design Considerations

In designing an equipment to display the capabilities of the multi-aperture ferrite cores, the magnitude of the driving and output pulses must be taken into consideration especially as those on the output may well have an amplitude of the order of 2½ V with a duration of 1½ µs. Thus, for the visual display of such output pulses a good quality oscilloscope is necessary. One pulse would hardly show up even on a fast time-base and, if it did, there would be some doubt as to its original wave-shape.

This has led to the use of a continuous pulse-forming circuit so that a visual display may easily be obtained. As stated in the introduction, output power may be obtained by using the minor aperture of each core as a transformer, the energy source being a winding through the minor aperture carrying a 10 kc/s sine wave. As the volume of magnetic material

around the minor aperture is small then the amount of power transferred is limited. The amount of power available was insufficient to light a bulb for the normal lighting of the laboratory.

The test equipment was developed to the following specification:

- (a) be a simple robust movable display;
- (b) show the capabilities of this shift register;
- (c) simulate various conditions in which the cores are required to operate to discover their limitations;
- (d) have good visual display.

Requirement (d) was probably the hardest to achieve since all the pulses involved are of short duration. The problem was successfully overcome by making the register a re-circulating one and using an oscilloscope. By putting information in the register and then re-circulating it and looking at the display on the oscilloscope, which was synchronized to the control, this requirement was achieved.

The overall block diagram of an equipment, designed initially for demonstration at the R.E.C.M.F. Exhibition in London in 1961, is shown in Fig. 4. Since this date further work has been carried out in order to incorporate a sophisticated control gear for the automatic control of a complete production line for explosives where the control equipment had always to fail-safe.

With reference to Fig. 4, the demonstration equipment can be broken down into four basic units.

- (a) a re-circulating shift register and drive unit,
- (b) control equipment,
- (c) power supply, and
- (d) display equipment.

To analyse the control circuit and its operation, the actual shift register must be analysed so that its working conditions are fully understood. The actual specification of the 10-bit shift register was found to be, by previous analysis:

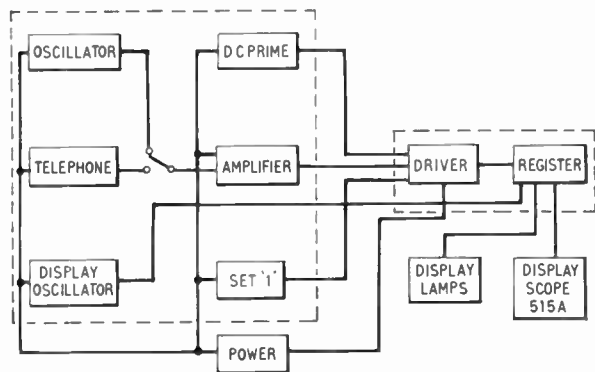


Fig. 4. Block diagram of the test and display equipment.

- (1) Advance pulse $2\frac{1}{2}$ A $\pm 10\%$. Rise time $1.6 \mu\text{s}$ $\pm 0.2 \mu\text{s}$; fall time $2.4 \mu\text{s}$ $\pm 0.4 \mu\text{s}$. Source greater than 1000 ohms (the advance pulse shape is shown in Fig. 6).
- (2) Prime current $100 \text{ mA} \pm 10\%$.
- (3) Set pulse 200 mA , $5 \mu\text{s}$ duration minimum.
- (4) Read-out, open circuit approximately 2.8 V for a minimum load resistance of 12 ohms .
- (5) 1 to 0 discrimination for read-out 8 to 1.

In order to start building equipment which will control the multi-aperture cores the shift register is the obvious starting point, since the performance of the cores is already known.

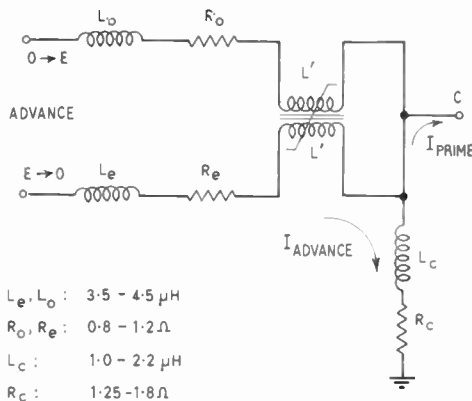


Fig. 5. Equivalent circuit of 10-bit shift register.

3. Shift Register

The shift register used General Ceramics' cores and had a capacity of 10 bits, using two cores per bit. The wiring of the register is as shown in Fig. 3. For the purpose of the design of the control equipment, the register may be broken down into a four-terminal passive network (Fig. 5), the circuit containing linear parameters L and R associated with the various windings on the cores. Non-linear elements L' appear in the odd and even windings and are coupled to each other, the value depending on the amount of information contained within the cores. Advance pulse currents for both the even and odd cores flow between terminals O and G and between E and G (not simultaneously). During an advance pulse, a much smaller d.c. prime current must flow between G and C . The external circuit must contain isolating elements so that the currents are contained within these paths. The non-linear element L' refers to the differing terminal characteristics when information is contained within the cores which may not be in the 'set' or 'prime' condition. When the four terminal network is being pulsed from the O circuit, the L' element produces back voltage proportional to the amount of information being shifted within the

register (similarly with the E element when it is pulsed). It is important that when the O circuit is being pulsed with the voltage created by this coupling L' , it has no effect on the E windings; hence, the E terminal must be open circuit while the O cores are being driven. The significance of the values given on this equivalent circuit will become apparent as the shape and magnitude of the advance current pulse, either applied to O or E , is applied to the register. The current and rate of change of current produce a back voltage in these elements, this back voltage setting an absolute minimum on the source voltage requirements of the driver circuit. In designing such a driver circuit that will in fact shift information from one core to the next, the temperature variations both due to ambient changes and heating of the cores resulting from these advance pulses, must be taken into consideration. The coercive force of a core changes by approximately 0.75% per deg C ; thus, if the cores are to work over a wide temperature range, the driver circuit must be capable of varying the current pulse amplitude in order to cover the range.

The actual pulse shape required to transmit information from one core to the next is shown in Fig. 6. This shape has been worked out for the particular cores in use and the type of windings involved. It is known that the register is primarily sensitive to the rise di/dt of the pulse, the decay time being less critical and the register being insensitive to amplitude above a minimum value. In this particular circuit energy will be developed as heat and it is essential to keep the amplitude to a minimum to avoid over-heating the cores. Simply, this means that all cores are set in the required condition for transmitting information down the register, a d.c. prime being continuously run through the transfer minor apertures. The variation of this d.c. prime will in fact control the successful operation of the shift register.

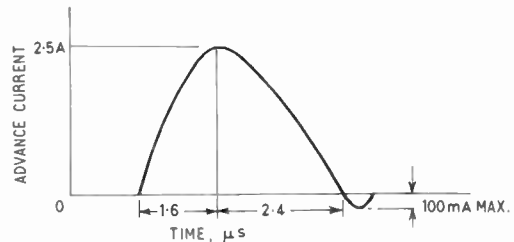


Fig. 6. Advance pulse.

4. Driver Circuit

In order to transfer information, it is essential to have a drive pulse which will look into the equivalent circuit of the shift register and give a pulse with the required rise-time to allow shifting of information from one core to the next.

This drive pulse, having an amplitude of approximately 2½ A, requires a large 'switch' with the necessary controlling trigger circuitry. The switch part of this equipment is in fact a thyatron. The circuit for this particular pulse generator is shown in Fig. 7 and has been chosen for the following reasons:

(a) Thyratrons are capable of handling peak currents of the amplitude required to drive the register. The particular thyatron used (type 2D21) can handle up to 10A pulse current without damage to the valve.

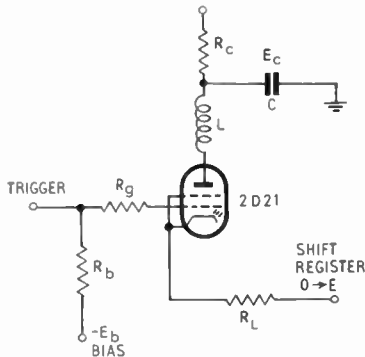


Fig. 7. Drive circuit for one side only repeated for $E \rightarrow 0$ cores.

(b) The circuit is relatively cheap, the trigger circuits being small, and the number of components can be kept to a minimum.

(c) The circuit is reliable; the only disadvantage of such a pulse forming circuit is the frequency of operation which is determined by the de-ionizing time and de-ionizing current of the thyatron. As the shift register is only required to shift at speeds of 10 kc/s or less, this is no disadvantage for this particular equipment.

Before a trigger pulse is applied, the thyatron is held in a low conductance mode by the bias E_b and the capacitor C charges through R_c . Thus, the voltage across C will be $E_c = (1 - e^{-t/R_c C})$. At some time T_1 , after C begins charging, a trigger voltage is applied to the grid of the thyatron which conducts.

With reference to Fig. 8 the equivalent circuit of the discharge path of the driver and shift register is shown, this being only half the circuit as the other shift period is identical, one pulse being used for the even cores and the other for the odd cores. If the register contains no information within the cores then the equations of the currents I_c and I_p will be:

$$E = \frac{1}{C} \int (I_c + I_p) dt + I_c R_c \quad \dots\dots(1)$$

and

$$0 = L \frac{dI_p}{dt} + RI_p + \frac{1}{C} \int (I_c + I_p) dt \quad \dots\dots(2)$$

The analysis of these two equations is shown in the Appendix, the results being:

(a) Capacitor voltage during charging

$$E_c = E(1 - e^{-t/R_c C}) \quad (\text{for } E_{c(t=0)} = 0) \quad \dots\dots(19)$$

$$E_c = E + (E_{c(0)} - E)e^{-t/R_c} \quad (\text{for } E_{c(t=0)} \neq 0) \quad \dots\dots(20)$$

(b) Pulse current

$$I_p = \frac{E_c}{\omega L} e^{-at} \sin \omega t \quad \dots\dots(21)$$

valid for

$$\frac{RE}{R_c E_c} \ll 1; \quad \frac{R^2}{4L^2} \neq \frac{1}{LC};$$

$$\frac{E}{R_c} \approx 0; \quad RR_c \gg \frac{L}{C}; \quad R_c \gg R$$

(c) Peak pulse current

$$I_{\text{peak}} = \frac{E}{\omega L} e^{-at_r} \sin \omega t_r$$

(Same conditions as (b) above).

(d) Rise time

$$t_r = \frac{1}{\omega} \tan^{-1} \frac{\omega}{a} \quad \dots\dots(22)$$

(e) Fall time

$$t_f = \frac{\pi}{\omega} - t_r \quad \dots\dots(23)$$

To check the above equations a driver was built with the following circuit parameters:

- $L = 62 \times 10^{-6}$ henries
- $C = 0.02 \times 10^{-6}$ farads
- $R = 28$ ohms
- $R_c = 10 \times 10^3$ ohms
- $E = 200$ volts
- $R_c = 20 \times 10^3$ ohms
- $R_b = 100 \times 10^3$ ohms
- $E_b = 3-20$ volts

These gave the calculated and measured values:

	Calculated	Measured
t_r	1.75 μ s	1.7 μ s
t_f	2.3 μ s	2.6 μ s
I_{peak}	2.78 A	2.6 A

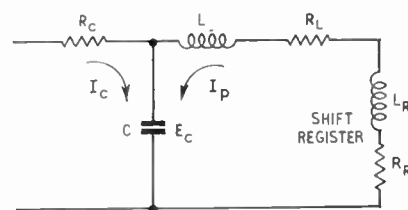


Fig. 8. Circuit for discharge of capacitor C.

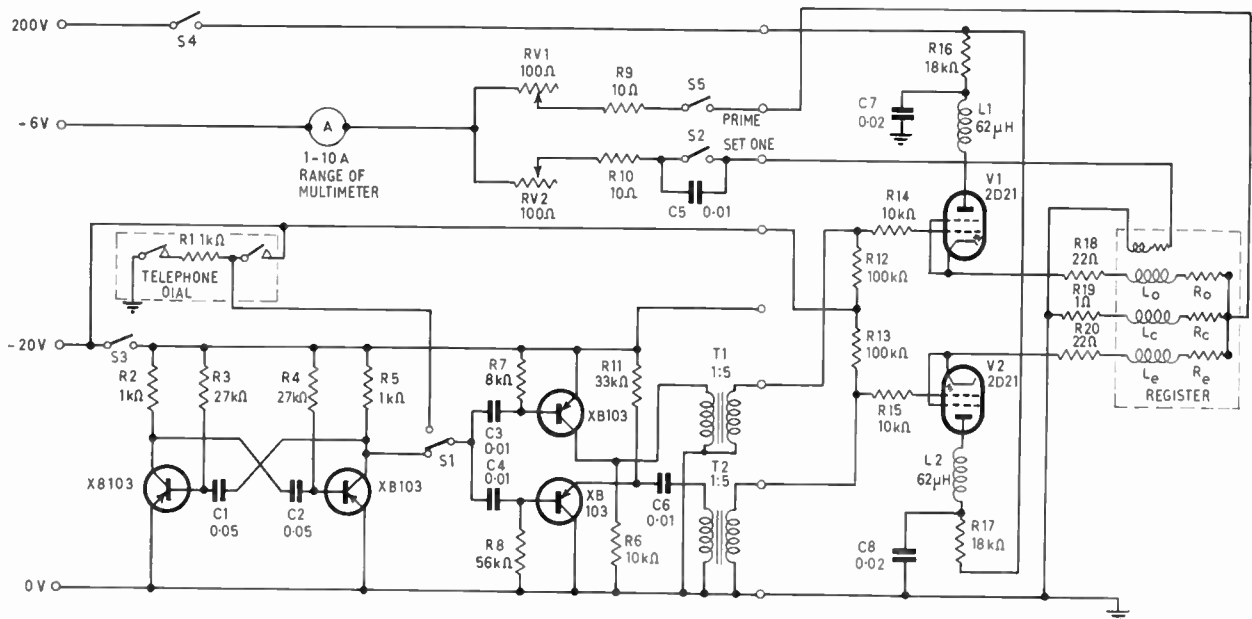


Fig. 9. Circuit diagram of the equipment.

Assuming that the register contains no information, all the derived expressions are valid under these conditions. It is found that when the register contains information in each core the pulse shape is not unduly affected and a 10% decrease in current is noted with this particular type of driver. The actual pulse amplitude required to drive a thyatron into the conducting mode is found to be 60 V with a source impedance of 1500 ohms, the pulse having a rise-time of 1 μ s and duration of 8 μ s. All these parameters can be varied by approximately 15% without altering the correct functions of the driven circuit. The output pulses obtained are within the specified rise and fall times, namely 1.6 and 2.4 μ s \pm 0.4 μ s. The maximum frequency at which the driver will work satisfactorily is 21 kc/s which is far beyond the capabilities of the particular cores used in the shift registers since the ferrite cores are basically 'lazy' and require time for priming and switching. In this particular instance the frequency of operation is not critical and is kept below 10 kc/s.

5. Control Circuit

The circuit for the control of the thyatron driver prime voltage and set pulse is shown in Fig. 9. This unit consists of a telephone dial, a square-wave oscillator, two amplifiers, a set prime potentiometer, and a set pulse potentiometer with the five switches, S1, telephone continuous drive; S2, set one; S3, transistor power; S4, thyatron power; S5, d.c. prime. The telephone dial is incorporated in this piece of equipment so that pulses can be sent to the driver to shift information down the register, thus

controlling the sequence by which set information can be put into the core. Since the shift register is a re-circulating one, any information put in the first core will continue to be driven around the register if the square wave oscillator is feeding pulses into the drive circuit. By means of the switch S1, pulses from the telephone dial or continuous pulses from the oscillator can be fed through the amplifier into the thyatron driver to drive the register.

The control for feeding information into the first core of the register is the switch S2, the current being controlled by a 100 Ω potentiometer and a 10 Ω overload resistor. The flux around the minor aperture must be reversed if information is to be contained within the core. The ampere-turns must be sufficient to change all the flux otherwise an analogue loading of the core would result. The capacitor across the switch S2 insures that no spurious pulses are fed to the core on breaking of the switch contact. The 'd.c. prime' circuit is similar to the 'set one' circuit and both circuits are connected to a multimeter which will indicate the current being taken by both circuits. Since both the prime and the set circuits can be controlled, the 'd.c. prime' and 'set pulse' parameters of the register can be ascertained.

The simple transistor square-wave oscillator is a free-running multivibrator circuit. The frequency of operation of this oscillator is determined by the 27 k Ω resistors R3 and R4 and the 0.05 μ F capacitors C1 and C2. The pulse repetition frequency is governed by the time constant $C_1 \cdot R_3 (= C_2 R_4)$.

Since pulses are required alternately for the O and E driver, one amplifier has been designed with a

grounded emitter configuration and the other has been designed with a grounded collector configuration, thus giving the necessary phase change required to trigger the thyratrons. The amplifiers are so designed that they are insensitive to spurious pulses which may be produced by the telephone dial. Both amplifiers feed trigger pulses through transformers each with a 1 : 5 voltage ratio so that the circuit is 'isolated' and the pulses from the amplifiers are of the correct amplitude, namely 60 V, to trigger the grids on the thyratrons. A capacitor C6 (0.01 μ F) is connected in series with the primary winding of the output transformer of the grounded collector amplifier so that the d.c. path from the 20 V supply to earth is broken. Switch S3 controls the operation of the oscillator and the switch S4 controls the 250 V to the thyratrons. Control of the drive circuit, the d.c. prime and the set pulse is thus obtained and provides the basic equipment for controlling the operation of the shift register.

6. Setting Up

The procedure for setting up the equipment for demonstration and evaluation of the cores has been kept as simple as possible. A 1- Ω resistor R19 (Fig. 9) is incorporated in the return line of the driving network. This resistor is included to monitor the pulses, which are present in the cores, from each half of the driving circuit. It is important when setting-up this equipment that the drive pulses into both the 'even' and 'odd' cores are checked by connecting the oscilloscope across R19. The pulses should be as shown in Fig. 6; the parameters are derived from the particular settings of the oscilloscope, taking into account any attenuators which may be used in the oscilloscope leads. Having checked that the correct pulses are present, the prime current is set. This current should be 100 mA and can be set by the potentiometer RV1. The pulse which feeds information into the first core of the shift register can also be checked at the same time by varying the potentiometer RV2 to give a reading of 200 mA on the meter. The oscilloscope is then connected across one of the output windings on the final core of the register and the d.c. prime is switched off momentarily. Provided the switch S1 is in the 'continuous operation' position

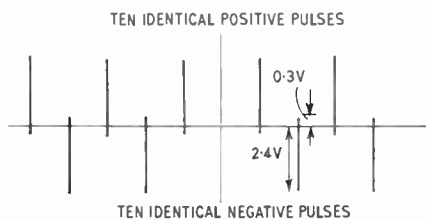


Fig. 10. Pulses from minor output core with register full.

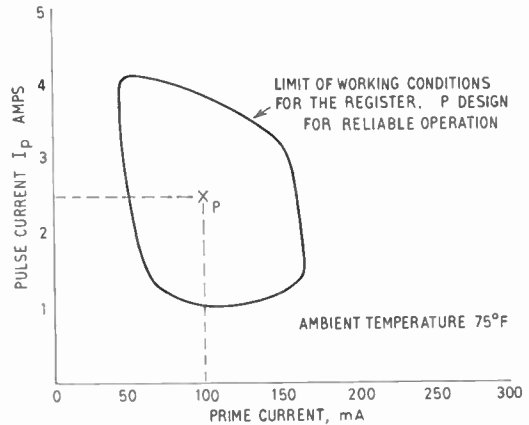


Fig. 11. Range map.

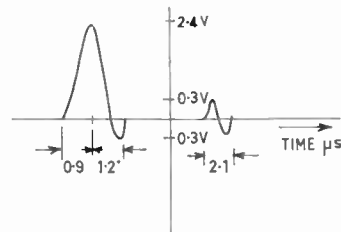


Fig. 12. Discrimination on 'one' and 'zero' in the minor core output.

all information contained within the core will be destroyed. The switch S2 of the 'set one' pulse is momentarily closed and this will fill the register with information. Provided the oscilloscope is set for a reading of about 4 V and 8 kc/s, then the wave shape should be as shown in Fig. 10.

Information is switched into the first core by means of S2 whilst the control equipment is in 'continuous operation'. The register is then filled with information as the recirculating time is shorter than the time taken to put information into the first core by closure of the switch S2. The information can be destroyed by shorting out any of the windings on the output of the cores or by switching off the d.c. prime. The limitations of the register can be evaluated by varying the d.c. bias current and the 'set one' current. The current pulse is adjusted by varying C₇ and C₈ between 0.03 μ F and 0.01 μ F. By varying these parameters, a range map can be drawn from the limits at which the prime of pulse currents operate the register successfully. The range map for the particular cores used is shown in Fig. 11.

Visual display of the output pulses from the cores during switching, shows a wave shape which is similar to that of the drive pulse, being first positive and then negative as the core is switched. Figure 12 shows this pulse and the pulse which appears when no

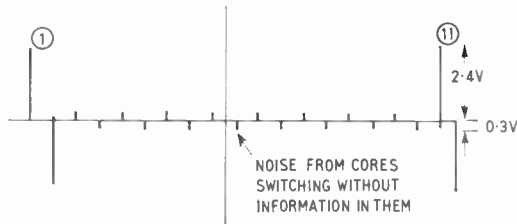


Fig. 13. Pulses from minor output core with 'one' in register.

information is contained within the core. The open circuit voltage from the minor aperture output windings, when information is contained, is approximately 2.4 V and when no information is in the core it is approximately 0.3 V. This gives a discrimination ratio of 8 : 1.

7. Demonstrations

It can be shown that information can be fed into the shift register in any desired sequence. This is achieved by first clearing the cores of information by switching off the d.c. prime. The switch S1 is then switched to 'telephone' and a '1' is put into the first core by momentarily closing S2. The switch S1 is then turned back to 'continuous' and the trace on the oscilloscope, which is connected to the output winding on the register, should be similar to that shown in Fig. 13. This indicates that the register is filled with information in the order 10000000001 and the time scale on the oscilloscope is arranged so that the first and the eleventh pulses are visible. The information is again destroyed and by using the telephone a sequence of information can then be put into the register and displayed to show its characteristics as a non-destructive shift register (non-destructive referring to the information contained within the cores). With normal single aperture cores the information contained within them is destroyed when

reading out. This condition does not apply in multi-aperture cores as the information is contained within the minor aperture area of the core. The information in the core will not be destroyed unless an advance pulse is passed through the major aperture.

The non-destructive characteristics are demonstrated by filling the register with pre-determined information, disconnecting all the supplies and then re-connecting them again. On the re-connection of the supplies it is seen that the information has not been destroyed and is still capable of being read out.

8. 'Fail-Safe' Control Unit

Having built this test equipment, it can be seen that the cores could be put to practical advantage in a 'fail-safe' batch production unit, e.g. one for use in the automation of the production of explosives.

The overall block diagram of such a unit is shown in Fig. 14. The specification for this is such that the shift register can be used as a means of detecting faults in a continuous automatic production line. It is important that the 'fail-safe' conditions are built into the control equipment as certain of the 'go/no-go' tests may be of a dangerous nature. The original piece of equipment was made up of a chain of interlocking relays, but since each relay had to fail-safe, this involved a considerable number of relays and interlocks. By using a shift register in this particular unit, only those relays which control the d.c. prime running through each individual core were used, hence, saving space and power and increasing reliability.

This 'fail-safe' control unit has a 10-bit shift register, wired as in Fig. 3 except for one change. All the 'odd' core d.c. prime leads are brought out and wired to 10 relays having normally open contacts.

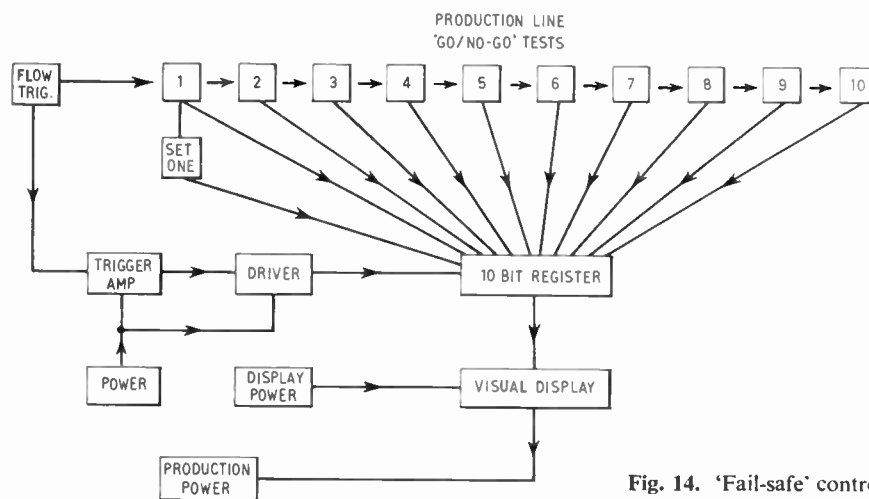


Fig. 14. 'Fail-safe' control unit.

The driver, trigger amplifier, and power supply are made up similarly to those of the test equipment. The visual display in this case was a separately powered 10 kc/s oscillator wired through the minor apertures. Small bulbs were powered from each even core, each bulb being suitably masked so that it was visible under normal ambient lighting. A trigger pulse from the output of each of the even cores holds a bi-stable flip-flop circuit which, when triggered, controls the power supplying the production line.

When the power is first turned on and the components to be tested are passed in front of a photo cell a trigger pulse is fed into the register via the trigger amplifier. Each time a component passes the photo cell, the trigger amplifier is actuated. Since a '1' is fed into the register during the first 'go/no-go' test, and provided all the tests numbered 1 to 10 are completed, the relays will close to provide the d.c. prime path through each core, and a '1' is automatically fed in the shift register. The equipment will give a visual display as each lamp will be lit at all times; pulses will keep the bi-stable flip-flops in operation, hence the main power to the production line will not fail. If a relay fails to close, the production line will stop but the visual display will be unaffected as it has its own power supply. A quick visual check on the number of lamps lit will indicate where the production line failed. By previously knowing which stage is dangerous it can be ascertained if it is safe for an operator to repair the fault in the equipment, or the particular fault in the component on the production line. Any component that fails will fail in the safe condition even if the main grid supply to the area is interrupted. If one of the bulbs fails, the equipment will automatically stop.

9. Conclusions

The design of test and display equipment for a demonstration at the 1961 R.E.C.M.F. exhibition has led to the development of an interesting and successful control unit. Its overall safety factor on the production line is made greater by the use of ferrite cores as the basis of the control mechanism. In applications not calling for fast operation this type of equipment has advantages over those using transistors or relay systems.

10. References and Bibliography

1. P. Broomer, "Multi-aperture devices", *Radio Electronic Components*, 2, No. 9, pp. 645-8, September 1961.
2. Hewitt D. Crane, "A high-speed logic system using magnetic elements and connecting wire only", *Proc. Inst. Radio Engrs*, 47, pp. 63-73, January 1959.
3. D. R. Bennion, H. D. Crane, "Design and analysis of MAD transfer circuitry", 1959 *Proceedings of the Western Joint Computer Conference*, pp. 21-36.

4. J. A. Rajchman and H. D. Crane, "Current steering in magnetic circuits", *Trans. Inst. Radio Engrs (Electronic Computers)*, EC-6, pp. 21-30, March 1957.
5. H. W. Abbott and J. J. Suran, "Multihole ferrite core configurations and applications", *Proc. Inst. Radio Engrs*, 45, pp. 1081-93, August 1957.
6. A. Wang and W. D. Woo, "Static magnetic storage and delay line", *J. Appl. Phys.*, 21, pp. 49-54, January 1950.
7. E. A. Sands, "Behaviour of rectangular hysteresis loop magnetic materials under current pulse conditions", *Proc. Inst. Radio Engrs*, 40, pp. 1246-50, October 1952.
8. M. Karnaugh, "Pulse-switching circuits using magnetic cores", *Proc. Inst. Radio Engrs*, 43, pp. 570-83; May, 1955.
9. S. Cuterman, R. D. Kodis and S. Ruhman, "Logical and control functions performed with magnetic cores", *Proc. Inst. Radio Engrs*, 43, pp. 291-8; March 1955.
10. Chester H. Page, "Physical Mathematics", (Van Nostrand, New York, 1955).

11. Appendix 1

Derivation of Driver Circuit Equations

$$E = \frac{1}{C} \int (I_c + I_p) dt + I_c R_c \quad \dots\dots(1)$$

$$0 = L \frac{dI_p}{dt} + RI_p + \frac{1}{C} \int (I_c + I_p) dt \quad \dots\dots(2)$$

where $R = R_R + R_L \quad L = L_R + L_L$

Transforming and solving for the transform $I_p(S)$

From (1) $\frac{\bar{E} - \bar{E}_c}{S} = \bar{I}_c \left(R_c + \frac{1}{S} \right) + \bar{I}_p \frac{1}{S_c} \quad \dots\dots(3)$

and (2) becomes

$$- \frac{\bar{E}_c}{S} = \bar{I}_c \frac{1}{S_c} + I_p \left(R + SL + \frac{1}{S_c} \right) \quad \dots\dots(4)$$

Eliminate I_c from (3) and (4)

Firstly multiply (4) by $(1 + SCR_c)$.

$$- \frac{E_c}{S} (1 + SCR_c) = I_c \left(R_c + \frac{1}{SC} \right) + I_p \left(R + SL + \frac{1}{SC} \right) (1 + SCR_c) \quad \dots\dots(5)$$

Subtract (5) from (3).

$$\frac{E + E_c SCR_c}{S} = I_p \frac{1}{SC} - I_p \left(R + SL + \frac{1}{SC} \right) (1 + SCR_c) \quad \dots\dots(6)$$

Expanding (6) and dividing by CR_cL

$$\frac{E}{CR_cL} + \frac{E_c S}{L} = -SI_p \left[S^2 + S \left(\frac{R}{L} + \frac{1}{CR_c} \right) + \frac{R_c + R}{R_c LC} \right] \quad \dots\dots(7)$$

$$I_{p(s)} = - \frac{S \frac{E_c}{L} + \frac{E}{CR_cL}}{S \left(S^2 + S \frac{R \cdot R_c + L/C}{R_cL} + \frac{R_c + R}{R_cLC} \right)} \dots\dots(8)$$

With $R_cR \gg L/C$ and $R_c \gg R$, eqn. (8) reduces to

$$I_{p(s)} = \frac{-S \frac{E_c}{L} + \frac{E}{R_cLC}}{S \left(S^2 + S \frac{R}{L} + \frac{1}{LC} \right)} \dots\dots(9)$$

If $b = \frac{E_c}{L}$; $m = \frac{E}{R_cLC}$; $a = \frac{R}{2L}$; $\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$

then eqn. (9) becomes

$$I_{p(s)} = \frac{-bs + m}{S((s+a)^2 + \omega^2)} \dots\dots(10)$$

Solution of (10) as a function of time.

$$I_p(t) = - \left[\frac{E}{R_c} + \left\{ \frac{E_c}{\omega L} \left(1 - \frac{RE}{R_cE_c} \right) \times \right. \right. \\ \left. \left. \times \sin \omega t - \frac{E}{R_c} \cos \omega t \right\} e^{-at} \right] \dots\dots(11)$$

An approximation to (11) which is valid for small values of $\frac{E}{R_c}$ and $\frac{RE}{R_cE_c} \ll 1$ provided $\frac{R^2}{4L^2} \neq \frac{1}{LC}$ is

$$I_p(t) \simeq - \frac{E_c}{\omega L} e^{-at} \sin \omega t \dots\dots(12)$$

To determine the rise time t_r of the current pulse (0–100%) the differential of (12) is equated to zero:

$$- \frac{E_c}{\omega L} e^{-at} \cos \omega t(\omega) + \frac{E_c}{\omega L} e^{-at} \sin \omega t(a) = 0 \dots\dots(13)$$

thus
$$t_r = \frac{1}{\omega} \tan^{-1} \frac{\omega}{a} \dots\dots(14)$$

Combining (12) and (14) the expression for the peak current is obtained:

$$I_{p(\text{peak})} = - \frac{E_c}{\omega L} e^{-at_r} \sin(\omega t_r) \dots\dots(15)$$

$$I_{p(\text{peak})} = - \frac{E_c}{\omega L} e^{-\left(\frac{a}{\omega} \tan^{-1} \frac{\omega}{a}\right)} \sin \left(\tan^{-1} \frac{\omega}{a} \right) \dots\dots(16)$$

Fall time at

$$I_p = 0 \text{ at } t = \frac{\pi}{\omega}$$

Thus fall time

$$t_f = \frac{\pi}{\omega} - \frac{1}{\omega} \tan^{-1} \frac{\omega}{a} \dots\dots(17)$$

$$t_f = \frac{\pi}{\omega} - t_r \dots\dots(18)$$

Manuscript first received by the Institution on 5th April 1962 and in final form on 14th October 1962. (Paper No. 857.)

Electronic Queueing Control for Materials Handling

By

GORDON CLARK, B.A.†

Presented at the Convention on "Electronics and Productivity" in Southampton on 19th April 1963.

Summary: Situations often occur in industry where several machines or operatives have a common source of material. If demands for more materials are dealt with in random sequence, or in accordance with a fixed routine, there will be cases where a machine or operator runs right out of material while others get their demands met more quickly than necessary.

A solution to this problem is to have an electronic system which acts like a queue and controls the routing of material so that demands are dealt with in the order in which they arise. Several versions of electronic queue are described together with their application in industry.

Problems which bear resemblance to queues also arise in connection with conveyors; it is sometimes necessary to identify packages as they are fed to a common conveyor, and then divert them at the appropriate collecting points along the route; sometimes inspection points cannot be located at the same place as the reject diversion mechanism, and a delay varying with conveyor speed must be introduced. It is shown that electronic queueing systems can be used with advantage in these situations also.

1. Introduction

'First come, first served', is a familiar principle which is widely accepted in human affairs, and generally a desirable situation in machine organization. However, whereas human bodies can be persuaded to stand in line, machines cannot. Hence, some mechanical or electronic analogue of the queue must be made, with the input to its 'tail' controlled by machine demand, and the output at its 'head' controlled by fulfilment of the machine demand.

There are several variations of the simple queue, and there are also distribution systems which bear a marked resemblance to a queue.

The essential feature of most of such systems is that 'labels' must be put into 'compartments' which are made to move along in such a way that the label reaches the output point or points at the correct time.

With distribution systems an actual label can often be physically attached to the item it describes, but where this is not possible a shift register can be used. This is an electronic equivalent of the compartments described above, the labels being represented by the state ('1' or '0', 'on' or 'off') of the flip-flops which constitute the compartments.

It is the aim of this paper to examine queueing and distribution systems, to show how these can be defined, and how practical electronic hardware can be built to control the inputs and outputs to such systems.

† Formerly Mullard Equipment Limited, Crawley, Sussex.

2. Fixed Length Queues

The simplest form of queue is the fixed length queue with a single entry point and a single exit point, i.e. one in which the number of items in the queue never changes. To define this queue only two parameters are needed, the number of items in the queue, called N and the possible types in the queue, called T . Industrially, such a system occurs where an article is inspected prior to entering a wrapping machine, and is then routed to the appropriate channel after the wrapping machine. If the inspection classifies the articles into six groups, and there are twenty articles in the wrapping machine between the inspection point and the channelling mechanism, the queue can be defined as a single simple entry $N_{20}T_6$ single exit queue.

Such a system can be made electronically by using a shift register as shown in Fig. 1. (For a detailed explanation of shift register see Appendix 1.) Figure 1(a) shows a system in which the six types are represented by six lines of the shift register. Figure 1(b) shows the same system modified to use binary code for the types. This gives a saving in the number of elements in the shift register, but introduces the need for an encoding unit at the entry point and a decoding unit at the exit.

In both systems the shift command is generated by signals from the conveyor or transfer mechanism. Photo-pick-ups, magnetic pick-ups, or contacts can be used, but the essential thing is that the shift pulses

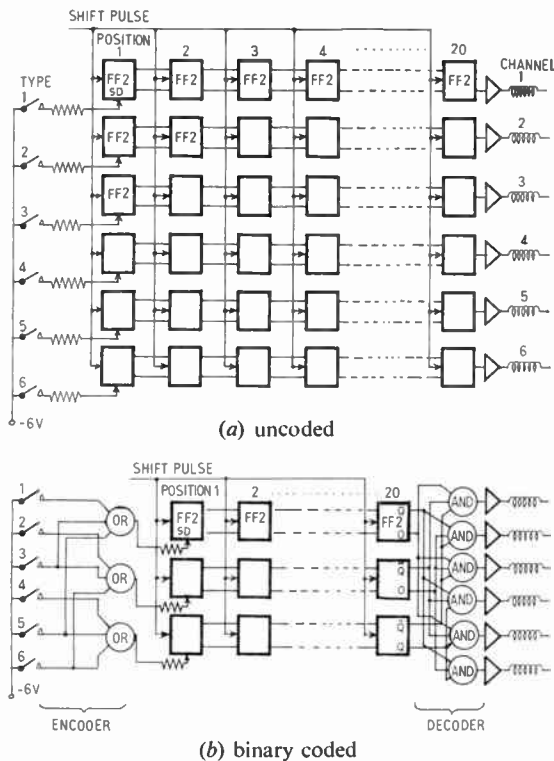


Fig. 1. Single-entry fixed-length single-exit queue

must be locked exactly to machine speed. Hence, it is not possible to use pulses separated by a fixed time interval, and if any belt drives are used it is essential that the pick-up is attached to the final moving element. In the case of articles which are not rigidly attached to the conveyor and hence may slip slightly, it is often most convenient to detect the passage of an article rather than the movement of the conveyor.

Where the slip between input and output varies by an amount approaching the interval between articles, or where articles are randomly spaced, the queue cannot be defined in such simple terms and the simple shift register approach is not satisfactory.

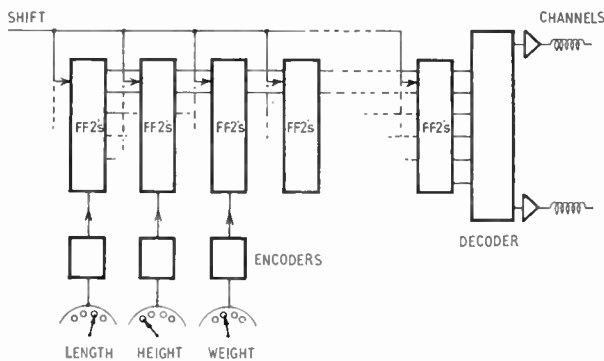


Fig. 2. Multiple entry.

A minor variation on such a queue is required where the inspection does not take place at a single point. For instance, a product might be inspected for length, height and weight at different stations and classified accordingly. In this case a multiple simple entry queue must be used as shown in Fig. 2. Again the shift register can carry binary coded or uncoded information. It is possible in such a scheme to overwrite existing information in the queue, and when designing a scheme of this sort economies can sometimes be effected by using this principle.

When dealing with inspection schemes we are concerned with labelling an article already on a conveyor. However, many handling schemes require

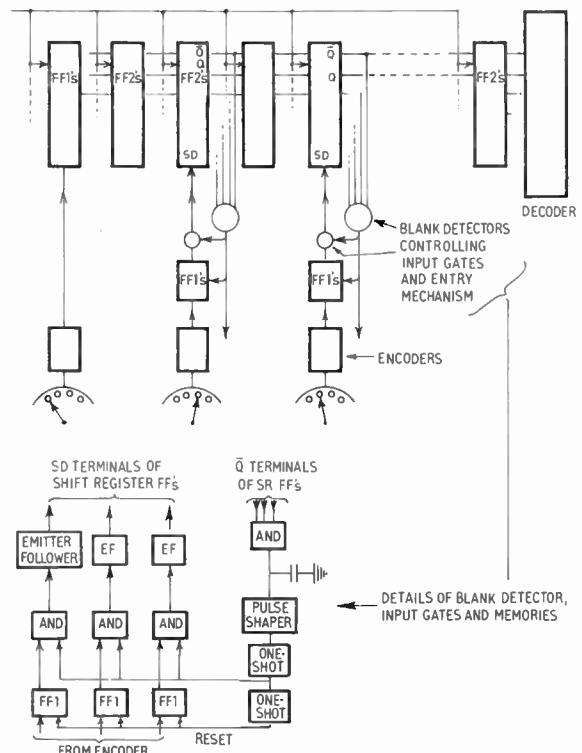


Fig. 3. Multiple conditional entry.

that an article shall be placed on a conveyor at points other than the start. Here the article and its label must wait until a space is available on the conveyor and in the electronic queue. The fact that the queue contains a blank is a convenient way of detecting when there is a space on the conveyor, and thus the article can be loaded on to the conveyor at the right time.

This type of queue can be called multiple conditional entry $N_n T_t$ single exit. Figure 3 shows such a scheme, and it can be seen that it is the same as Fig. 2 with the addition of the 'blank' detector. The

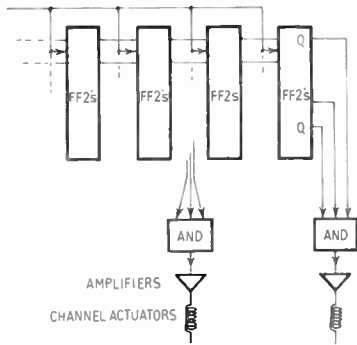


Fig. 4. Multiple exit.

scheme is shown with memories at each entry point, to store the label. Depending on the method of generating the label information these may not be needed.

The queues considered so far have only had a single servicing point—an article reaches the servicing point at the same time as its label reaches the end of the shift register, and is dealt with accordingly. However, a very common requirement is for articles with different labels to be diverted from the conveyor at different points. This type of queue can be classified as multiple exit, the other parameters being classified as described earlier.

Such a system is shown in Fig. 4, and it will be noted that a coded shift register is used, and consequently decoders are required for each exit position. However, since such a scheme usually requires only one category of article to be diverted at each exit point, the decoder consists of a simple AND gate, wired (or switched) to give an output when the required code appears in the shift register.

3. Variable Length Queues

All the schemes discussed so far have been concerned with handling articles moving in a well defined progression. In a more general queueing situation, the input to the tail of the queue is of a random nature, and is not synchronized in any way with the 'servicing' at the head of the queue. The most noticeable result of this difference is that the number of bodies in the queue (N) is now a variable number.

For many applications N will vary between 0 and T , where T is the number of different types, as before. However, cases can occur where N can be larger than T ; this implies that some items appear more than once in the queue.

A more common requirement is where N is limited to a number considerably less than T . In this case it is obviously possible to have a rush of inputs into the queue and hence exceed its capacity. Under these circumstances an alarm may be raised, or the items

which fail to get a place in the queue can be dealt with in random fashion.

There are four basic methods of coping with the variable queue:

- (a) variable entry point
- (b) variable exit point
- (c) 'first-up' circuits
- (d) automatic shunt.

3.1. Variable Entry Point

In a single variable entry $N_{0-4}T_{12}$ single exit queue the first item to be put in the empty queue must be inserted in the last place in the queue. The second item (assuming the first has not been dealt with) must be inserted in the third (penultimate) place, the third item must be inserted in the second place and the fourth item in the first place. Any subsequent items attempting to find places in the queue must raise the alarm.

If the first one is dealt with after 3 items have been put in the queue, items two and three must move up one place, and now the fourth item must be inserted into position two, not one. The point of entry therefore, must be controlled by a switching system which steps back one place for every new item added to the queue. Also, whenever an item is removed from the head of the queue, all the other items in the queue must advance one place.

Such a switching system could be made with electro-mechanical devices, but it can be made with solid state logical units. In this case a reversible counter determines the point of entry by opening gates at inputs to the various flip-flops which form the places in the queue.

Figure 5 shows the block diagram of a variable entry $N_{0-4}T_{12}$ queue, and the principle can be extended to much larger queues. However, there are limitations in such an arrangement as it stands since no precautions are taken against simultaneous inputs,

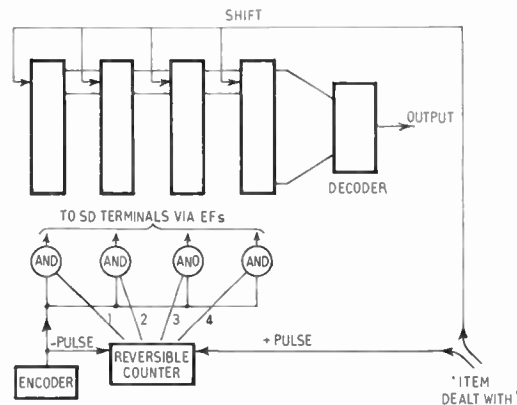


Fig. 5. Variable entry.

nor against the queue being moved up one place at the same time as a new item is being inserted.

3.2. Variable Exit Point

This method is very similar to the one just described, except that in this case the point of exit is varied rather than the point of entry. Here again a reversible counter is used to open the appropriate gates, and now it is stepped forwards by the arrival of a new item, and stepped backwards when an item has been dealt with (see Fig. 6). With this method the items in the queue are all moved forward one place by the arrival of new information, rather than by the servicing of the item at the head of the queue. This arrangement, like the one described in the previous section, cannot cope with different information presented simultaneously.

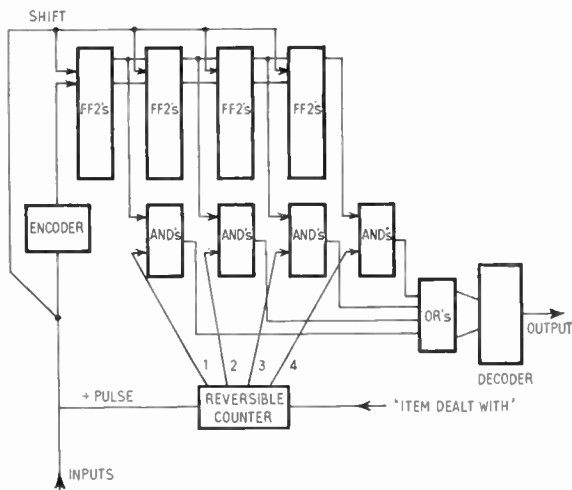


Fig. 6. Variable exit.

It should be noted that in both the variable entry and the variable exit systems there is no need to clear an item from the queue when it has been dealt with. In the case of the variable entry system the information stored in the flip-flops at the head of the queue is destroyed when a shift pulse is applied; with the variable exit system the information goes into the 'dead' part of the queue, until it is finally lost at the end.

3.3. First-up Circuits

Two 'OR-invert' (NOR) circuits† can be connected with the output of each feeding the other, and such an arrangement gives a store with two stable states, '10' and '01'. If, however, drive is applied to the other inputs of the NOR circuits, both inverters will give a '0' output (see Fig. 7), and now the first input going

† Appendix 3 gives a detailed explanation of NOR circuits as used in Norbits.

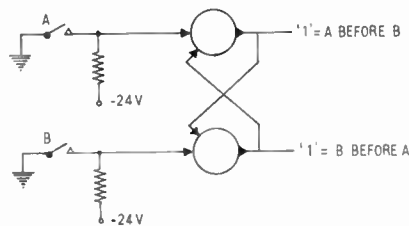


Fig. 7. Basic 'first up' circuit.

to '0' will give a '1' output. If subsequently the input to the second NOR goes to '0' this will not change the state of the store.

Thus, we can use the state of the store to determine which input changed to a '0' first, and hence we have a queue of 2 which will determine whether A is before B or whether B is before A. By using three NOR stores we can determine the priority of three items: A before or after B, B before or after C, C before or after A.

It is necessary to inspect two of the stores to determine the head of the queue, since only when A is before B and before C is it at the head of the queue. This inspection can be done very easily with NOR circuits, by looking at the other outputs of the memories which give a '0' output when A is first. The original input from A must also be fed into the inspecting NOR, since when using the inverted store output above we cannot determine that A is before B; we can only determine that B is not before A. Hence, we may get '0's fed into the inspecting NOR even before the A input operates (i.e. changes from '1' to '0').

In such a scheme as this the number of NOR circuits required is $T + 2 \sum_{i=1}^{T-1} (= T^2)$ where T is the number of types. The $2 \sum_{i=1}^{T-1}$ NORs are used for the 'first-up' memories and further T NORs are needed for the inspection. Where T is as large as or larger than the maximum number of inputs to a NOR circuit, the number of NORs used for inspecting must be increased, or diodes can be used to increase the 'fan-in'.

Although the number of NOR stores in this system increases rapidly as the number of types increases, it should be remembered that this is a full queue, i.e. the number of places in the queue is always equal to the number of types, which is not necessarily so for shift register queues.

When an item at the head has been dealt with the input from that channel must change from '0' back to '1'. The inspecting NORs then indicate the next in the queue. This is repeated until all items in the queue have been dealt with. Meanwhile, other items can be added. Figure 8 shows a queue for four items. This queue is particularly suited to the tending of

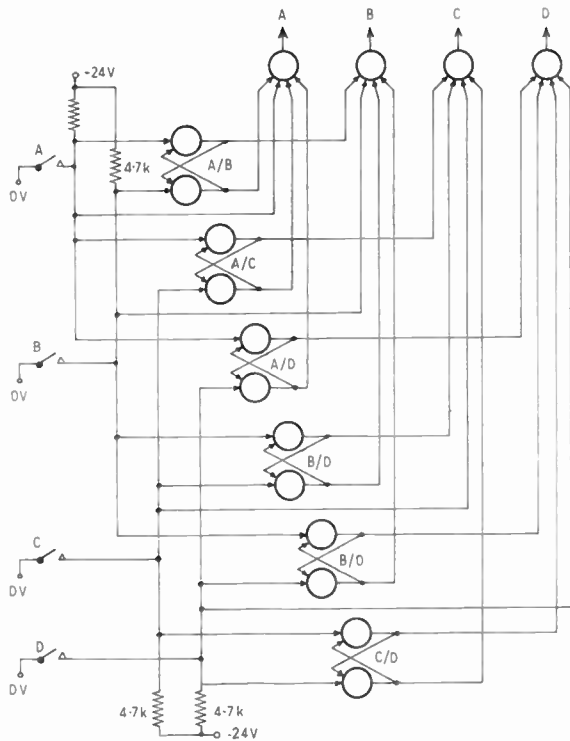


Fig. 8. Queue of four using 'first up' circuits.

machines, since it is easy to arrange that a machine needing attention provides a continuous signal, which disappears when it has received attention. Whereas the queues described earlier require that the input information is presented and then removed, this type of queue requires a continuous input.

By the very nature of the 'first-up' circuit it does not matter if two lots of input information are presented simultaneously. The store concerned with both inputs (there must always be one and only one such) will take up a '01' or '10' state arbitrarily, allocating priority to one item or the other. Similarly it does not matter that an input changes to the 'serviced' position (0→1) at the same time as another input changes to the 'demand' (1→0) position.

3.4. Automatic Shunt

In the 'automatic shunt' system information is inserted in the first place of the queue and then automatically moved up to the head of the queue by shift pulses which are generated continuously. The shift pulses to each place in the queue are inhibited if there is any information in the place in front of it. Thus successive items placed in the tail of the queue are shunted towards the head, and held stationary when they come behind the previous item. When an item is dealt with, the place at the head of the queue is emptied, and this leaves a space, so that succeeding items are once again shunted towards the head of the queue.

If normal shift register flip-flops are used to form the queue it is necessary to reset the flip-flops in each place as soon as the information has been transferred to the next place, since the shift operation does not destroy the source of information. Unless this resetting facility is included in the design the movement of an item down the queue will leave a trail of information behind it which will prevent succeeding items from closing up.

A simpler solution is to use counting flip-flops† and to apply the shift pulses to their reset terminals. Any flip-flops which are in the '1' state will generate a 'carry' on being reset, and this carry will set the succeeding flip-flop. Such a system is shown in Fig. 9 and it can be seen that AND gates are used to inhibit the shift pulses, by connecting the inputs to the Q (inverted) outputs of the flip-flops. Thus, the shift pulses only pass through the gate when all the flip-flops are at '0'.

The shift pulse is a negative pulse, and the back edge (a positive-going step) effects the shift operation. The use of this back edge technique prevents information being shifted more than one place at each shift pulse, and ensures that the information contained in the flip-flops of each position is moved in parallel down the queue.

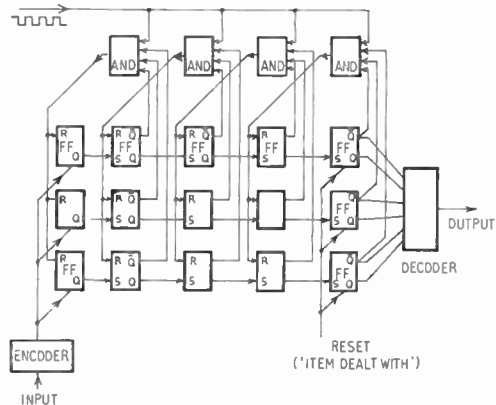


Fig. 9. Automatic shunt.

When a 'carry' is being fed into the S input of a flip-flop, it will often occur that the shift gate is open and hence a positive edge will be being fed to the R input. However, the S input will always set the flip-flop in spite of the R input, since the internal steering circuits of a counting flip-flop prevent the R input being effective if the flip-flop is in the '0' state; (the flip-flop must be in the '0' state when a 'carry' is produced by the previous flip-flop, since this '0' state is necessary to allow the shift (reset) pulse to be gated to the previous flip-flop).

† See Appendix 2.

The resetting of the flip-flops at the head of the queue is controlled by a circuit similar to that used for the inputs, in order to ensure that a full length shift pulse is allowed through the gate after resetting.

The use of such a strobing system limits the speed of operation of the queue and introduces delays in the acceptance of input information, but if 'Combi-elements'† are used a complete scan of a 40-channel input system can be effected in less than one millisecond, which is more than fast enough for the vast majority of industrial applications.

Figure 11(a) shows in block diagram form a single strobed entry $N_{0-4}T_{12}$ single exit queue. Figure 11(b) shows parts of the system in more detail.

5. Variable Speed Distribution Systems

Where the movement in a mechanical handling scheme is derived not from a conveyor but from motors built into the moving item or by gravity, different items on a distribution route may move at different speeds. Provided that a single entry single exit queue is satisfactory, normal variable length queueing techniques can be used. As each container enters the route its label is placed in a queue, and the label for the first item will be at the head of the queue, followed by the label for the second item and so on. The fact that the information may reach the head of the queue long before the item it describes reaches the end of the route is of no consequence, since the label remains at the head of the queue until its item has been dealt with.

A more common requirement is for a multiple exit system, and one approach to this problem is to have a simple shift register type of queue in which the shift pulses for each stage are generated by pick-ups located along the route at intervals corresponding to the minimum space between articles. This method would work but it has two disadvantages. First, the number of pick-ups used is large, and experience in industrial installations has shown that the installation and maintenance of large numbers of pick-ups is expensive and frequently a source of trouble. Second, the number of stages in the shift register is decided on the assumption that the route is filled completely with items which all have minimum spacing, whereas the rate of entry may be known to be much lower than this, and hence the minimum space only occurs when one item catches another up in transit; thus every small space produced must give rise to a corresponding larger space.

A better approach is to use variable length queueing techniques, with one queue for the section of route between the entry and the first exit, and further queues between exit points. The 'first-up' circuit

can be used, but the transfer of information from one queue to the next must use special a.c. coupled circuits, to avoid 'race' conditions. Each queue must be a full queue, i.e. it would cope with N_i items in each queue where there are T_i types; this is not required in most cases, so this facility is an unnecessary expense.

The 'automatic shunt' system can be very easily adapted to multiple exit requirements by using the passage of an item past an exit point to set the 'clear' flip-flop which generates the reset pulse to the last stage in the queue. The 'carries' produced by the last stage of one queue are fed into the first stage of the next queue, and then are automatically shunted in that queue.

Such a system is satisfactory only when self-driven containers on a route are cleared of their contents at their exit point and themselves allowed to proceed. If, however, the container is diverted complete the label will proceed on down the queue without any corresponding item. Thus the number of reset pulses fed into the last place of succeeding queues will not correspond with the number of labels in the queue, and the system will get out of step.

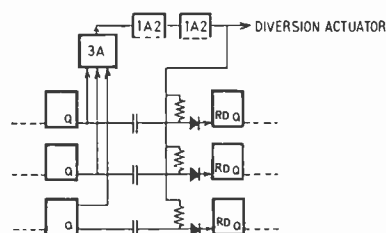


Fig. 12. Carry gates.

This situation can be dealt with by inhibiting the 'carries' when the code appropriate to an exit is at the head of the queue. A detector for the appropriate code (a decoder) is required in any case at each exit point in order to actuate the diversion mechanism, and the output of this decoder controls the carry inhibit gates. Figure 12 shows how two successive queues are linked by pulse logic units which gate the 'carries'.

As the various types are diverted at their exit points the T factor is reduced in successive queues, and it is possible to arrange the coding so that the queues can be tapered. However, the saving in equipment is liable to be offset by the increased design time and loss of flexibility.

6. Beam Ahead

A completely different approach to the variable speed distribution problem is the 'beam ahead' system. In this system the label is fed into the tail of the queue which is made with NOR stores interconnected so that

† Mullard Equipment Ltd.

each set of memories adopts the state of the previous set of stores. Thus information put in at the beginning of the queue appears at the end almost instantaneously.

In addition to the label queue there is a position queue, consisting of a single store for each section of the route. An article travelling along the route sets a store as it enters each section, and the output from that store resets the previous position store and also applies an inhibit signal to the information stores in the previous section. Thus, as an article travels along the route it creates a 'blank' in the section behind it.

New information fed in at the beginning of the information queue will now only propagate along the queue as far as the blank. It will be appreciated that as each article must occupy at least two positions in the queue, roughly double the amount of hardware is required as compared with the flip-flop shift register approach. However, the creation of the blank may be desirable in some applications, where for mechanical reasons it is necessary to maintain a minimum space between items, and this blank in the electronic queue can be used to prevent the movement of an article in that section of the route.

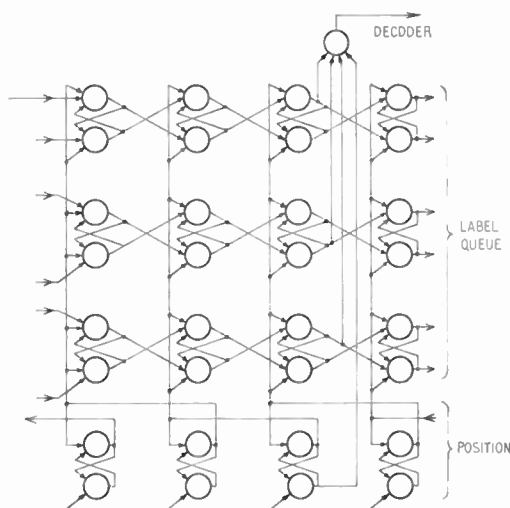


Fig. 13. Beam ahead.

A further advantage of this form of queue is that it is d.c. coupled, and hence heavy filtering can be used to eliminate pulse interference, which can be a serious problem in a.c.-coupled circuits. Figure 13 shows a distribution system designed on these principles.

7. Separating Position and Information

In the 'beam ahead' system information and position are dealt with in separate queues, and this principle can also sometimes be used to effect a saving in synchronous distribution systems, where the position

of an item on the conveyor must be defined in smaller increments than are required for information.

For instance, when dealing with articles of between 6 in and 12 in, with a minimum gap of 2 in between articles, it would be necessary to control the diversion mechanisms to open within 2 in of a desired position, whereas articles occupy a minimum of 8 in of conveyor space.

In this case the information queue would be driven by shift pulses generated at 8-in intervals, whereas the position register would be driven by pulses generated at 2-in intervals.

The diversion mechanism only operates when the appropriate information code is present as well as a '1' in the position register.

8. Mixed Systems

Various types of queuing and distribution systems have been described but no mention has been made of mixed systems, though cases can obviously occur where this is desirable.

Though it is technically feasible to link NOR logic systems with a.c. coupled shift register or counting systems this is generally not very satisfactory since there is a tendency to combine the disadvantage of both systems.

9. Conclusion

Various types of queuing and distribution systems have been described, and illustrated with designs using NORBITS and 100 kc/s circuit blocks (Combi-elements). However, the principles are universal and could be applied to other 'logic blocks' giving similar facilities.

The d.c. coupled systems using NORBITS are in general simpler to design and service, since there are no control pulses to generate, and no problems of pulse rise times. However, the shift register and automatic shunt types of queue and distribution system are much more economic in many applications, particularly where it is not necessary to queue all the possibilities.

10. Acknowledgments

The author wishes to acknowledge the work of Mr. J. Brown and Mr. P. Hamlin on queues, particularly the 'first-up' circuits, and to thank the Directors of Mullard Equipment Limited for permission to publish this paper.

11. Appendix 1: Shift Register Flip-Flop†

The circuit has various names, including the British Standard version, a bistable toggle with external control of pulse input gates. In the form illus-

† The Appendices are not intended to give comprehensive descriptions of 'Combi-elements' and 'Norbits', but only to explain their use in the particular cases described in the paper.

trated below (Fig. 14), the FF2, in the 100 kc/s circuit block ('Combi-element') range, the pulse input is a positive going edge applied to the 'P' terminal (since the input is via a capacitor the d.c. level is unimportant). From the input capacitors the pulse is fed to the bases of the two transistors via diodes, and these diodes are cut off by a '1' (-6 V) applied to the G_S and G_R terminals. When a '0' is applied to G_R and a '1' to G_S , the pulse is allowed to reach the right hand transistor, cutting it off, and prevented by the cut-off diode from reaching the left hand transistor. Thus the Q terminal is at -6 V '1' and the \bar{Q} at 0 V.

The values of the cross-coupling resistors are such that the circuit remains stable in this stage, until a new input stimulus is received.

If G_S is left at '1' and G_R at '0', subsequent pulse inputs to the P terminal will not change the state of the flip-flop. If however, the G_S is changed to '0' and G_R to '1' the next pulse into the P terminal will change the flip-flop, giving a '0' on Q and a '1' on \bar{Q} . It should be noted that the changing of G_S and G_R does not in itself affect the flip-flop; these inputs only 'steer' the next pulse input.

The pulse input capacitors ensure that a change at G_S and G_R only affects the potential on the two steering diodes after a delay of several microseconds. Thus if several FF2's are connected with the Q and \bar{Q} terminals of one feeding the G_S and G_R terminals of the next, and pulses are applied simultaneously to the P terminals of each, each flip-flop will adopt the state which the previous flip-flop was in before the 'shift' pulse.

A shift register so formed from FF2's can be of any length, and information (a '1' or '0') fed into the first

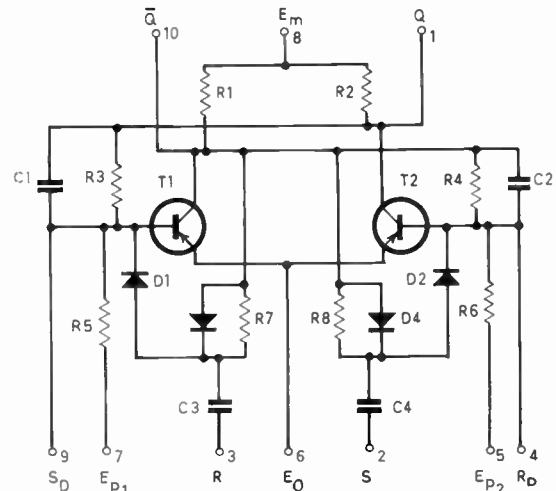


Fig. 15. Counting flip-flop.

flip-flop in the register will reach the last in the register after x pulses, where x is the number of flip-flops in the register.

Where the information to be carried in the register cannot be expressed simply as '1' or '0', several rows of flip-flops can be used, the pattern in the flip-flops at the beginning of the rows being transferred in parallel down each row, by shift pulses applied simultaneously to each flip-flop in every row.

The state of the flip-flops can be changed by applying signals direct to the bases of the transistors (S_D and R_D terminals), and this is often a more convenient way of inserting information than by using the G_S and G_R inputs.

12. Appendix 2: Counting Flip-Flop

This circuit (Fig. 15) is very similar to that of the shift register flip-flop, the main difference being that the 'steering' circuits, which are controlled by the G_S and G_R terminals on the FF2, are connected internally to the collectors of the transistors. The part of the circuit equivalent to the G_R point is connected to the Q terminal, and the equivalent of G_S to \bar{Q} .

Thus if an FF1 has a '1' on Q and a '0' on \bar{Q} , a positive going pulse applied to the S and R terminals (equivalent to 'P') will cause the flip-flop to change state.

In its most common application the FF1 is used with its S and R terminals linked, and it then functions as a binary divider; successive positive edges fed in cause the Q terminal to go first to -6 V and then back to 0 V, thus giving a positive edge output for every two in. Such circuits can be cascaded to form divider chains for any power of 2, and dividers for other numbers can be made by gating the carries and other such devices.

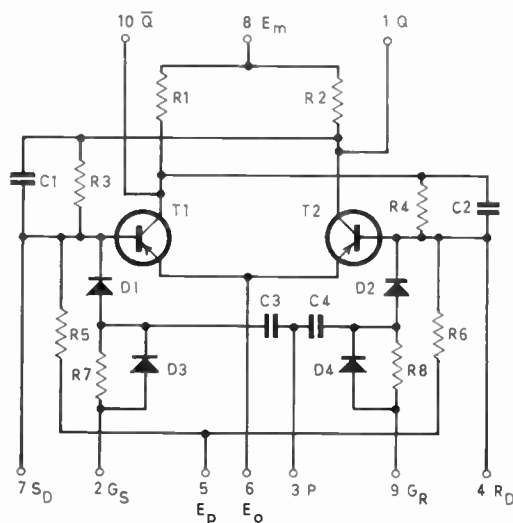


Fig. 14. Shift register flip-flop.

The FF1 circuit suffers from the disadvantage when used in the 'automatic shunt' queue that it gives a positive 'kick' out of its S terminal when it is reset by a pulse on its R terminal. This is very liable to affect the previous flip-flop, if its Q output is in the '1' state, and hence this 'kick' must be suppressed. A resistor and diode effectively do this, while still allowing the 'carry' to be propagated without hindrance.

Alternatively the new FF3 circuit could be used. This circuit has no 'kick-back' effects, and has the further advantage that its input thresholds have been raised, thus rendering it less subject to interference.

13. Appendix 3: The Norbit

The NOR circuit is used with many minor variations, and the NORBIT is a version particularly intended for relatively low-speed (1000 operations/second) industrial applications (Fig. 16).

With the NOR circuit, the transistor is normally held cut-off by the bias resistor connected to the +24 V line, and hence gives an unloaded output of -24 V. If any of the input resistors are connected to -24 V, (or another NOR with its transistor cut-off) the transistor conducts, and the output goes to 0 V.

If we call the negative level a '1', and the 0 V level '0', we can say that the output is a '1' for '0's on all its inputs, and a '0' for '1's on any of its inputs.

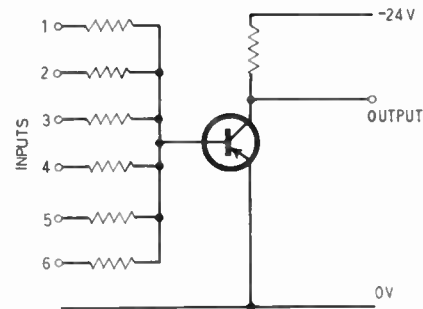


Fig. 16. Norbit.

Thus we have an AND function for '0's and an 'OR' function for '1's. Since the NOR circuit also performs the inverting function, we can make a cascade of AND-OR-AND-OR gates with NORBITS linked directly to each other.

The storage function can be produced by connecting two NORBITS so that the output of each feeds the input of the other. This makes a bistable circuit which can be changed by feeding a '1' into one of the unused inputs of the NORBITS.

Manuscript first received by the Institution on 11th March 1963. (Paper No. 858.)

© The British Institution of Radio Engineers, 1963

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10} , from nominal frequency for
September 1963

1963 September	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.	1963 September	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	—	-129.7	-32	16	-130.0	-131.7	-25
2	—	-131.4	-32	17	-131.3	-132.7	-24
3	-131.5	-131.2	-34	18	-131.7	-131.0	-24
4	-130.8	-132.0	-31	19	-130.8	-130.8	-24
5	-132.3	-131.9	-29	20	-130.1	-130.4	-22
6	-131.6	-131.6	-30	21	-129.9	-129.8	-21
7	-131.6	-131.2	-28	22	-129.7	-130.4	-21
8	-131.7	-131.3	-28	23	—	-130.9	-21
9	-131.2	-131.5	-28	24	-130.8	-130.5	-21
10	-131.2	-131.0	-27	25	-130.5	-131.1	-20
11	-131.0	-130.2	-27	26	-130.5	-131.4	-18
12	-130.9	-131.1	-27	27	-131.0	-129.5	-18
13	-130.0	-129.8	-26	28	-129.2	-129.6	-16
14	-130.3	—	-24	29	-129.1	-130.3	-16
15	-129.6	—	-24	30	-130.1	-129.8	-16

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium $F_{1,m}(4,0)-F_{1,m}(3,0)$ transition at zero magnetic field.