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 AUGUST 1963 NUMBER 2

REGISTRATION OF ENGINEERS

For the protection of the public there is in many countries legislation regulating the registration of persons practising particular professions and stipulating that such registration requires evidence of appropriate training and experience.

In Great Britain the legal profession first established such registration over 200 years ago, culminating in the Royal Charter granted to the Law Society in 1831. Other examples are the General Medical Council, registration with which is compulsory for all medical practitioners, and the Architects' Registration Council, originally sponsored by the Royal Institute of British Architects. Legislation has also been enacted to cover the Accountancy profession.

In this technological age there is increasing enquiry as to why there should not be registration for engineers. In the older branches of the engineering profession it has slowly become accepted that for public and other responsible appointments preference shall be given to corporate members of, for example, the Institution of Civil Engineers. Specious argument is often advanced that because the entire engineering profession contains many specialities, it is impossible to base the registration of engineers on adequate training and experience. Speciality in technology is, in fact, no more diffused than, for example, in the legal or medical professions. Indeed, in several countries, the professional description of 'engineer' is restricted to persons holding university or equivalent qualifications.

It is true that on basic standards of academic competence the engineering Institutions in Great Britain have been tardy in agreeing minimum requirements, although there is now a large measure of agreement between the Chartered Engineering Institutions on common standards. Immediately there is every prospect of agreement on a common Part I examination between most members of the Engineering Institutions Joint Council which was formed in 1962.

Radio and electronic engineers are entitled to some indication of their own Institution's policy in such matters. There has not, in fact, been any change in the policy expressed by an officer of the Institution twenty-three years ago. Sir Louis Sterling† then urged that the Institution should co-operate with all other bodies in establishing the status of the engineer. This was a theme that he developed in his Presidential Address in 1942‡ when, after referring to the impetus given to the development of radio and electronic science by the demands of war, he stated:—

> "This has thrown a responsibility on the radio and teaching professions. In this process of development there are possibilities for improving our national engineering educational system, for it must be remembered that technical training for our profession has been developed over a period of years in different localities and by various agencies—and often only in response to the immediate needs of groups of students with heterogeneous background and purpose.

† *J.Brit.I.R.E.*, **1** (New Series) No. 4, page 100, September 1940.
‡ *J.Brit.I.R.E.*, **3** (New Series) No. 2, page 33, September 1942.

"These wartime schemes must not be dropped; we should adhere to the principle of training engineers and, having trained them, ensure methods of registration. If registration is essential in war, it is equally necessary in peace, in order to ensure proper utilization of manpower.

"The idea of peace-time registration of engineers is by no means new. A few years ago the late Professor W. Cramp, addressing the British Association,† advocated the setting up of a body with statutory powers to define the qualifications of persons entitled to call themselves 'engineers'. Then the aim was to prevent unskilled persons from jeopardising life and to check unprofessional conduct; on this score alone registration is highly desirable, but it becomes even more desirable if also used for ensuring proper utilization of manpower. Since engineering is now recognised as so much a part of any nation's life and in fact, international life, it is obviously of some importance to consider what standards of professional conduct (and ability) are required of its practitioners.

"It may be argued that one of the main purposes of a 'Law for the Registration of Engineers' is to institute standards of competence which should be regarded as minimum qualifications for those wishing to be termed 'professional engineers'. It may be suggested in return that responsible organizations and Government Departments would only engage competent men, but that does not necessarily follow.

"For example, we have the many alterations and amendments made in the compilation of the wartime Central Register because, although excellent in principle, the Register did not at first take into consideration the specialist aspects of each and every profession—especially in the engineering industries. Consequently there was an almost fatal tendency to generalize instead of specialize. This was to a large extent quickly remedied, especially by the three Services, by encouraging co-operation from the professional engineering Institutions—the specialist bodies.

"If registration is desirable, then should it be undertaken by one central Government Department or by specialist bodies, such as our leading professional Institutions? Membership of a professional body not only necessitates educational and practical ability, but has the further advantage of entailing adherence to a professional code calling for a high standard of conduct."

Registration of engineers has also been debated in other Institutions‡ and there have been many references to Sir Louis Sterling's Presidential Address in technical journals, but no practical approach toward securing registration of engineers became possible until the major engineering Institutions were brought together by the formation of the Engineering Institutions Joint Council.§

The purpose of the E.I.J.C. and the goodwill which has attended its formation, would seem to justify the claim that it should be the body for the registration of engineers. This will first require that the E.I.J.C. itself should immediately seek legal status and ultimately a Royal Charter of Incorporation. In order to succeed in this purpose the fullest co-operation will be required from the existing Chartered Engineering Institutions. Without such help the E.I.J.C. must fail and with it, hope for several decades of the engineering profession speaking, as a profession, with a unified voice.

In considering the functions of E.I.J.C. it must be emphasized that it does *not* duplicate the purposes of any 'learned society'. Such activity is confined to the sponsoring bodies, but no learned society is admitted to membership of the E.I.J.C. unless it also has professional standards of competence for membership. The sponsoring Institutions have a combined membership of over 200,000 and together can promote the necessary legislation to obtain the registration of engineers and procure for the engineering profession the same status as that enjoyed by other professions.

G. D. C.

† "The Engineer and the Nation"—Address to the British Association, 1936. [Dr. C. C. Garrard, a Member and later President of the Institution, attended this meeting and associated himself with Professor Cramp's recommendation.]

‡ For example, J. W. Thomas, *J. Instn Elect. Engrs*, **82**, page 49, 1938.

§ *J.Brit.I.R.E.*, **25**, No. 6, page 478, June 1963.

U.D.C. 621–52: 681.32

Some Principles and Circuit Techniques for Controlling Machine Tools from a Central Digital Computer

By

D. A. BELL, M.A., Ph.D.†

AND

P. M. WILL, Ph.D.†

Presented at the Convention on "Electronics and Productivity" in Southampton on 18th April 1963.

Summary: The advantages and economics of time sharing a computer between a number of machine tool control systems are considered. Some of the circuit techniques necessary to achieve this are described, namely encoders, quantizers, function generators, priority sharing methods, and a matrix for parallel selection.

1. Towards the Computer-Controlled Factory

Shortly after the war the idea of a pushbutton factory was publicized, largely through the work of Norbert Wiener.[1, 2] Practical progress was slow, but in 1956 there was a further resurgence of interest in automation when the D.S.I.R. produced a report,[3] R. H. MacMillan's book "Automation: Friend or Foe?" was published,[4] and the enthusiasts began to talk about "sweeping the last man out of the factory".[5]

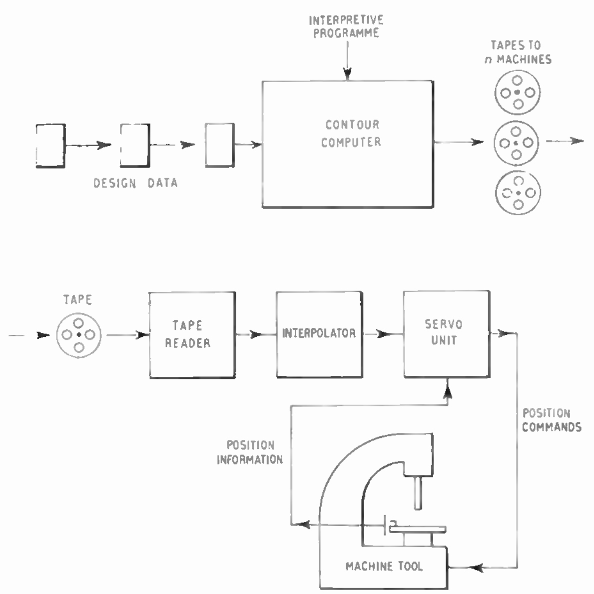


Fig. 1. General arrangement of machine tool control by a single computer employing magnetic tape.

By this time computers appeared to be an essential part of large-scale automation, and in 1958 Wilkes[6] suggested the use of a single high-speed computer to control a number of automatic devices. In one sense this had already been achieved in the Ferranti and E.M.I. systems of digitally-controlled milling machines (Fig. 1), for one computer is likely to be able to supply control tapes for several independent machines. However, the example involves only indirect control by the computer through the medium of a tape, and the computer enthusiast wants on-line control.

By 1962 on-line control by computers began to be a reality in three fields, namely oil refining processes, chemical process control and large electrical power stations. In the oil refining application there are still doubts whether it is economic to apply computer control to a single distillation or cracking process, but it is suggested that a computer could do valuable service in optimizing a whole refinery installation.[7]

In chemical process control it was first proposed to use a computer as an overall organizing and optimizing device to adjust the set points of local controllers maintaining desired values of flow, temperature, etc. But more recently[8] it has been suggested that the transfer of the differentiating and integrating functions of conventional 3-term controllers from the latter to a control computer will give improvements in economy and in reliability.

The application to power stations is three-fold. Firstly, in the starting of a conventional power station there are many operations which must be performed in the correct sequence and subject to checks on the completion of other operations or availability of supplies. This is basically a problem in logical decisions and as such is immediately suited to a computer, which can also provide data-logging and monitoring services. A second application is to nuclear power stations where the consequence of a catastrophic failure could be much more serious than for a conventional power station, and in some respects automatic controllers are less liable to error than human controllers. In addition there is a large data logging function, e.g. checking temperature readings on 1000 or more thermo-couples, which is much better performed by an electronic device than by a human being; and there are complex inter-relationships between the controlling variable and reactor response, which are readily handled by a computer. A third electric-power application of computers is to the

† AMF British Research Laboratory, Blounts Court, Reading, Berkshire.

calculation of load despatching, which is in effect the sharing of load between different power stations so as to obtain the optimum economic effect. In countries where there is more than one power supply company it is necessary to keep a record of the money value of power exchanges between the different networks, and this can readily be done by the same electronic computer which controls the occurrence of such exchanges of power.

Computers are also used in an organizing role in other contexts such as the control of a paper mill,[9] the charging of blast furnaces,[10] and the control of steel-making[11] and of steel mills.[12, 13]

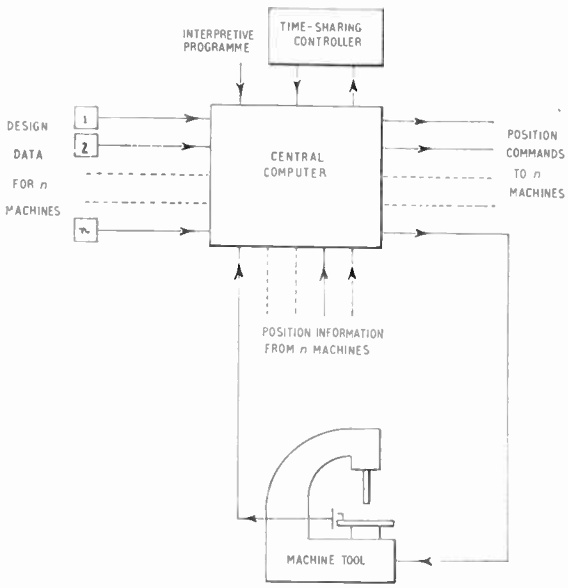


Fig. 2. Control of machine tools by a central computer using time-sharing techniques.

It can be said that in all the applications outlined above the processes are changing so slowly that speed of response is not a critical factor. The question now is whether there is a case for using a digital computer in the automatic factory for direct control of mechanical devices such as machine tools, in which changes occur over much shorter time intervals. It is often said in computer circles that the most economic way of handling a given amount of calculation is to concentrate it in the largest and fastest computer available, since the cost of a computer increases less rapidly than its computing power, and therefore all cybernetic activity should be concentrated in a central machine (Fig. 2). From an unsophisticated view-point, however, it seems more reasonable to carry out simple control operations by means of simple devices attached to the machine.

The first step towards computer control is to put the information in digital form. It has already been established that measurement and control signals should be quantized in order to secure easy repeatability, for this is the principle involved in all numerically-controlled tools whether they are of the co-ordinate-finding type only (e.g. jig borers) or continuous following devices (e.g. milling machines and flame cutters). Many existing machines use a pulse-count system in which the signal is quantized but is not coded, and therefore positions are found by dead-reckoning over a greater or lesser period from a datum point. The advantages of proceeding from a quantized signal to a binary-coded signal are that one has always an absolute value which is not dependent on having received correctly all past sequences, and that the information is more readily stored. An application requiring storage was an automatic machine for balancing crank shafts reported by Csech at a Paris Colloquium.[14] The figure for the weight of material to be removed was stored in binary form, since analogue storage could not readily be made accurate over the time between the determination of the unbalance of the shaft by means of a balancing machine and the application of a milling cutter to successive webs along the shaft to eliminate the unbalance.

By analogy with the case of storage it may be that quantized function-generators will prove more accurate (in the sense of retaining their functional characteristic over a long period) than analogue function generators. For the more complex function it may in any case be feasible to use a digital computer. One example of this is in the A.P.T. system, for programming by computer the operation of a milling machine to generate a three-dimensional surface. In this case the computer calculates the optimum line of cut and the distance for which it can be allowed to extend without exceeding the permitted tolerance of surface shape.

The idea of choosing cutting directions can be arrived at by a different argument. In setting up the positions of valves in a process controller, a computer is controlling *rates* (e.g. of flow) and the analogous action in a milling machine would be to control velocity, rather than position, the case in favour of which is as follows. Suppose the machine is cutting at 6 in/min and the tolerance is $\pm 0{\cdot}001$ in. Then in the worst case it will take 10 ms for the machine to run out to the limit of tolerance. But suppose the section of trace immediately ahead of the cutter is an inch of straight line. Then provided the components of velocity are initially correct and remain unaltered, the machine can wait 10 seconds for its next instruction. When cutting curves the periods will be shorter, but never as short as the 10 ms of the crude theory. If necessary, the periods can be kept long by reducing feed speeds when the radius of curvature is becoming small. The computer will, of course, compute velocity settings in terms of the difference between the next desired position and the present attained position, so that movement is always in the correct direction and there is no tendency for errors of position to accumulate. It may be objected that in this mode of control a continuous curve will be approached as the envelope of a number of tangents. But in fact a pulse-count method of position control will result in a curve being approximated by a set of steps, since no action will be taken until the difference between the actual and demanded values of one of the co-ordinates differs by the equivalent of one pulse. Both forms of trajectory may be modified by machine dynamics, but basically the envelope of tangents is at least as good an approximation as the average of steps. In either case some anticipation of future information is desirable, and this will be most readily obtained in a computer. If we now think of an average period between changes of 1 second and a computing time of up to 10 ms (compare the 300 μs required for setting all the parameters of a 3-term process controller[8]) it is possible for one computer to control up to 100 movements or 33 three-axis machines. The total cost of equipment on a present-day numerically controlled machine is difficult to apportion between position control, tape reader and interpreter (whether for magnetic or for paper tape) and share of the control computer. But if we estimate the cost of information-handling equipment at £2000 per controlled axis, we are allowed up to £200,000 for a computer which will do this work for 100 axes. The ratio of computing speed to cost is improving—Table 1 compares typical parameters of scientific-type computers of 1956 and 1963—and the feasibility of computer control is to that extent drawing nearer.

Table 1

	1956 Computer	1963 Computer
Mode	serial	parallel
Addition time	300 μs	1 μs
Multiplication time	2000 μs	14 μs
Direct-access store	55 single-word registers	32 768 words storage
Price (order of magnitude)	£40 000	£150 000

No complete computer-controlled installation is yet available, but the remainder of the paper will describe some of the circuit techniques which have been prepared for incorporation in such a system.

2. A Parallel Gray to Binary Code Converter using Transistors and NOR Logic

As remarked above, many of the existing schemes of numerical control of machine-tools depend on a pulse-count indication of position. Digital encoders are now becoming more common, and the avoidance of ambiguities in reading necessitates using a code pattern which is not the common binary code. The alternative codes are Gray or cyclic permuted binary, unit distance binary coded decimal, any of the unit distance codes described, for example, by Susskind[15] or the V-scannable codes described by Lippel.[16]

If such coding is used merely in a simple control loop, subtraction of the present position from the commanded position can be effected by special circuits, operating directly on Gray code.[17] But for general operations, particularly via a digital computer, the Gray code must be translated to common binary.

The characteristics of Gray code are well known and translation is usually effected in an explicitly serial manner,[18] using shift registers and clock pulses, and can then occur only at discrete times. Asynchronous translation is less well known, but can be carried out as follows:

(a) The most significant bit is unchanged and forms the first bit of the required binary number.

(b) This first binary bit is subtracted without carry from the second bit in the Gray number to give the second binary bit.

(c) The $(n-1)$th binary bit is subtracted without carry from the nth Gray bit to give the nth binary bit.

As an example, the Gray code number

0 1 1 1 0 0 1 0 1 1 0

becomes the binary number

0 1 0 1 1 1 0 0 1 0 0

Successive subtraction without carry requires a truth table equivalent to the negation of the logical

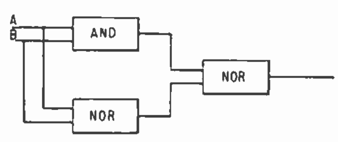


Fig. 3. Schematic of NOR logic circuit.

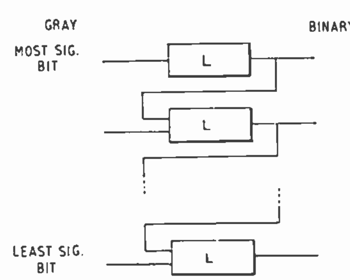


Fig. 4. Asynchronous converter.

statement A = B, which by de Morgan's rule may be replaced by $A \cdot B+\overline{A+B}$.

This statement may be realized by NOR logic in a straightforward manner, giving the schematic circuit of Fig. 3. The overall flow diagram of an asynchronous converter is shown in Fig. 4, where L stands for the logic circuit of Fig. 3. This circuit formed the basis of an experimental five-bit parallel decoder which had a propagation time for a change in the most significant bit of less than a microsecond. A further advantage of the asynchronous translator is that if a change occurs in the nth most significant Gray bit this can only affect binary bits $\geqslant n$ without the necessity for reconverting the whole number: this can be important in many applications.

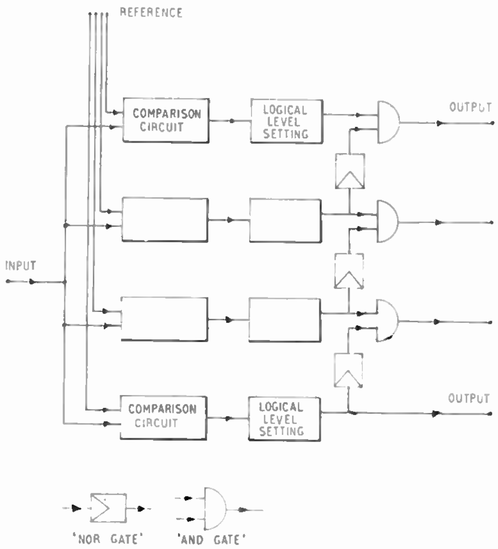


Fig. 5. Quantizer logical diagram.

3. A Quantizer with Parallel Output

Mechanical position can be represented in a binary digital form by combining a suitable encoder with a code translation device such as was described in the previous chapter. Some quantities, however, are difficult to obtain in other than analogue form (speed, acceleration, temperature, pressure, etc.) and the first step towards digitizing them is to quantize them. In the case where the quantum widths are equal the quantizer could be a standard analogue to digital converter.[19] In the case where the quantum widths need not be equal and in addition where few quantization levels are necessary, a simpler scheme is desirable.

The logic diagram of a simple quantizer in which the few quantization levels may be set arbitrarily, is shown in Fig. 5. The principle of operation is that if there exists a chain of voltage selection or comparison circuits which switch ON when the input voltage exceeds a set reference level, then the outputs of the first m of these will be ON when the input exceeds the mth reference level and does not exceed the $(m+1)$th reference level. This is sufficient information to perform the logical switching.

The mth signal and all lower signals must be switched OFF when the input exceeds the $(m+1)$th reference level. This logic may be implemented by a NOR gate driven from the $(m+1)$th level driving an AND gate in the mth line as shown in Fig. 5. For the sake of increased reliability in the case of failure of the comparison circuits, the NOR gate could be driven from all higher signals rather than from the single $(m+1)$th signal.

Levels may be selected by biased diodes, but the practical difficulty lies in the lack of sharpness of cut-off of the diodes.[20] Schmitt triggers are another possibility, but suffer from hysteresis. In the interests of reasonable precision, comparison with amplification was required; and for simplicity it was performed on the long-tailed pair amplitude discriminator of Fig. 6 in which the input was applied to the base of one transistor of the long-tailed pair and the bias voltage at which the circuit was to turn on was applied to the base of the other. Since the output of the discriminator was required to drive the logic circuits for line selection some power amplification was necessary.

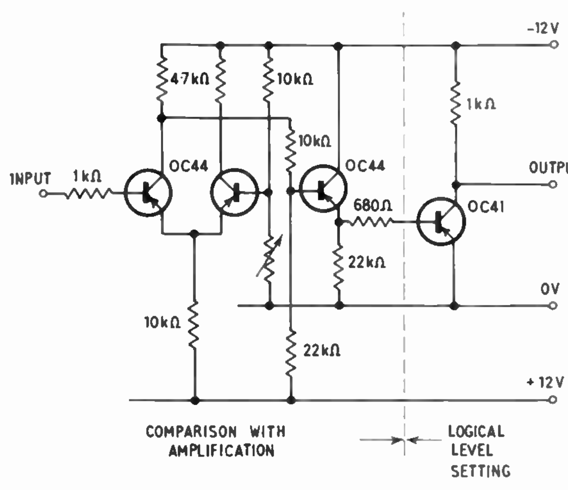


Fig. 6. Comparison and level setting circuit.

To this end the output of the differential amplifier was taken single-ended through an emitter follower, the emitter resistance of which was taken to earth while the base was biased positively when the input level was less than the required switching level and negatively when the input was greater than switching level. A single transistor inverter using an OC41 switching transistor was driven from the emitter follower, the low output impedance of which ensured that the OC41 was driven hard on and hard off giving logic levels of 0 V and −12 V. These levels give definite control of the NOR and AND gates.

The instrument described quantizes an input voltage into discrete levels which may be preset individually giving variable quantization width. The quantal information is contained in the presence or absence of a standard signal (−12 V) on one of a number of output channels equal to the number of comparison amplifiers. A particular channel is energized for the duration of time that the input signal remains in the defined quantum width: sampling of the input signal is not required.

As an example, a 4-level quantizer was constructed in the manner described, with bias levels of 0·5 V, 1·0 V, 2 V, and 4 V. When tested with sine and equilateral-triangular waves of up to 500 c/s, the output wires were in fact energized for the length of time in which the input signal fell within the prescribed upper and lower limit for each wire.

4. Digital-to-Analogue Function Generator

The advantages one may hope to gain from the use of absolute digital control of a system (as distinct from incremental-pulse control) include compatibility of input and output signals with computers and digital data-processors, and in particular the ability to perform complex calculations on a digital computer in order to modify the control parameters as required.

This possibility is limited to slow modifications (but see the timing of calculations suggested in the Appendix) so that either analogue or hybrid digital/analogue devices may be needed when mathematical operations have to be performed on the control signals themselves. This Section describes a digital-to-analogue function generator which can be used for these purposes and for linearizing signal flow paths and determining convenient performance criteria.

Binary-digital to analogue conversion may be accomplished by summing currents generated from the digits of the input number in resistors weighted according to a scale 2^{-n}. If the highest accuracy is desired, the circuit can be compensated against impedance changes.[21]

The first suggestion for the direct generation of functions of a binary input is a result of the question: what is the effect of weighting the summing resistors in other than the standard form 2^{-n}?

In detail it is noted that the critical points for a binary input are those corresponding to 2^n, i.e. a single 1 in the binary number, since it is characteristic of this system that during the intervals $2^n - 2^{n-1}$ of the binary input the entire response up to 2^{n-1} is repeated. Prescribing the law at the *points* 2^n prescribes the *entire* function in such a way that we have no control over the function at intermediate points.

To obviate this difficulty the next logical step is to change the resistance weighting with input signal magnitude.

The simplest scheme for accomplishing this change is shown in Fig. 7. The first function which was

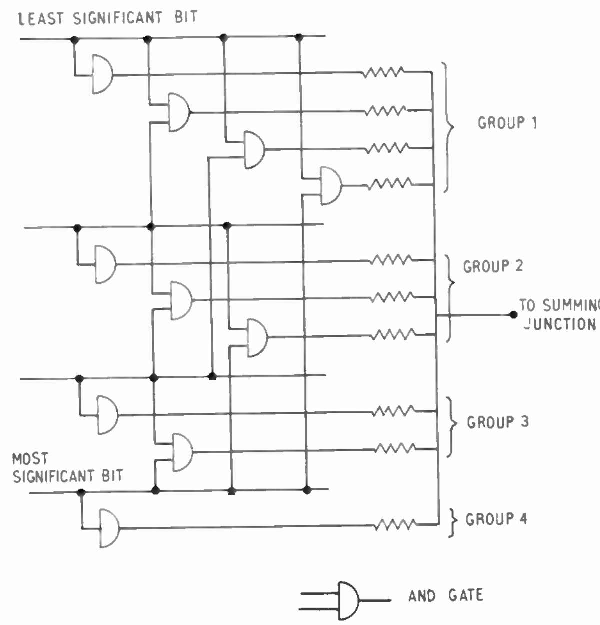


Fig. 7. Simplest digital to analogue function generator.

Group 1	Group 2	Group 3	Group 4
1 MΩ	470 kΩ	220 kΩ	100 kΩ
470 kΩ	220 kΩ	100 kΩ	
220 kΩ	100 kΩ		
100 kΩ			

generated by this method was the square. To give a square it is noted that the slope is a linear function of x and that the increase in slope is constant. Thus the resistive loads were chosen in pure binary progressions both within groups and from group to group.

The experimental response of the function generator was approximately a square law, but with large irregularities, because this circuit is composed of AND gates exclusively, so lower couplings affect the output throughout. When each higher digit is ON it switches into the summing mesh a current due to itself and currents from the AND gates from all lower digits. In addition extra current is obtained from the cross coupling of the lower digits themselves. These stray unwanted couplings give rise to the inaccuracies of the function. The cross-coupling may be compensated by adjusting the resistance values individually instead of taking the values calculated without cross-coupling. Figure 8 shows a square approximated by this method.

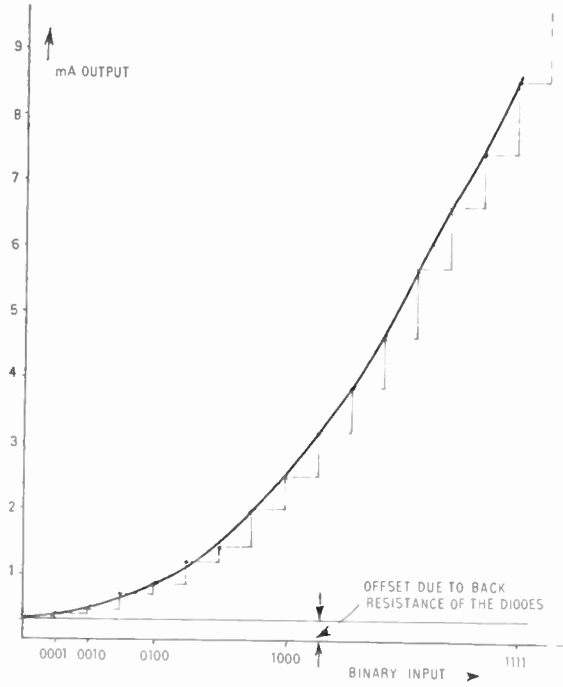


Fig. 8(*a*). Approximation to a square curve.

Even considering the offset at the origin due to the low-back-resistance germanium diodes used in the AND gates, the curve is within 2% of the full scale of the square function.

The curve of Fig. 8(*a*) was constructed by fitting the standard square curve starting from the origin. By fitting the curve from the right-hand end, i.e. by considering the digits $8-15$, $4-8$, $2-4$ then $1-2$, it is possible to take some account of leakage resistance and reduce the zero offset still further.

The AND gate circuit is satisfactory for functions having monotonically increasing slope, but it cannot produce functions of decreasing slope because the occurrence of the least significant digit in combination with others will always produce a slope at least as great as that at the origin. This difficulty may be overcome by including a sufficient number of NOR gates so that, for example, the resistor corresponding to the least significant digit is ON if and only if all higher digits are OFF.

This form of function generation is closely analogous to the analogue computer technique of choosing break points and fitting slopes to the required curve. In this case the break points are constrained to fit the critical points 2^n. Using a decoding tree it would be possible in principle to choose other break points but this was not pursued. Re-entrant functions are not capable of simulation by this method.

5. Orthogonal Generation and Multi-Variable Functions

In the system of Fig. 7, ten logical gates were used to generate 16 points on the function, and some cross-coupling was therefore inevitable. If there were 16 mutually independent summing resistors for the 16 points, the generation of the function would be described as an 'orthogonal' process. If a quantizer of the type described in Section 3 were used, with an output level for each point on the curve and an appropriate summing resistor associated with each output the desired independent setting of points would be attained. Decreasing and re-entrant functions can also be generated by this method.

This orthogonal method can be extended to the generation of functions of two variables, an operation which is complicated with ordinary analogue function generators. Referring to Fig. 9, a quantizer is used for each variable and the output lines arranged in lattice form as shown. The logic is that if the most significant line of quantizer 1 is ON then the variation of the input to quantizer 2 causes the gates of row 1 of the matrix of switches to operate. The summation of currents from these switches may then be caused to follow any prescribed law by independently set points: similarly if any other line is energized. (Fig. 10.)

This method will thus give a set of n^2 planes normal to the z axis as an approximation to the required sur-

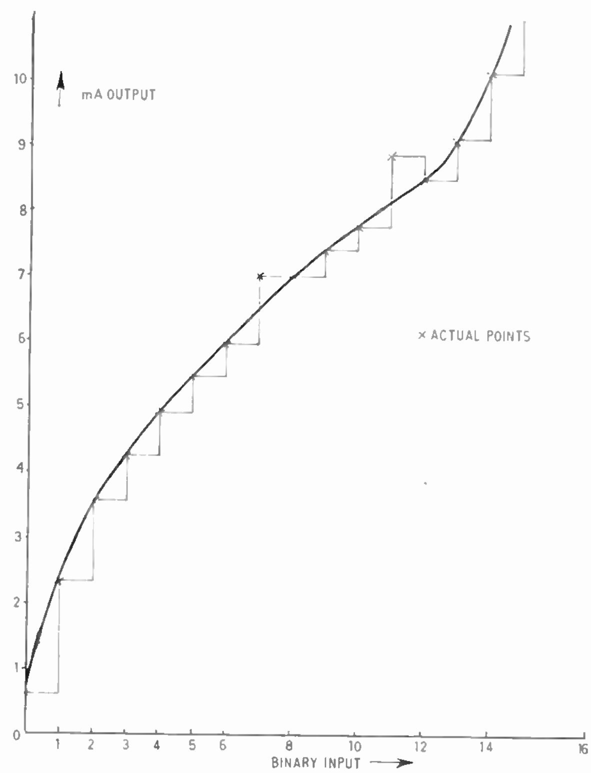


Fig. 8(*b*). Approximation to a square root (using a modification of Fig. 7).

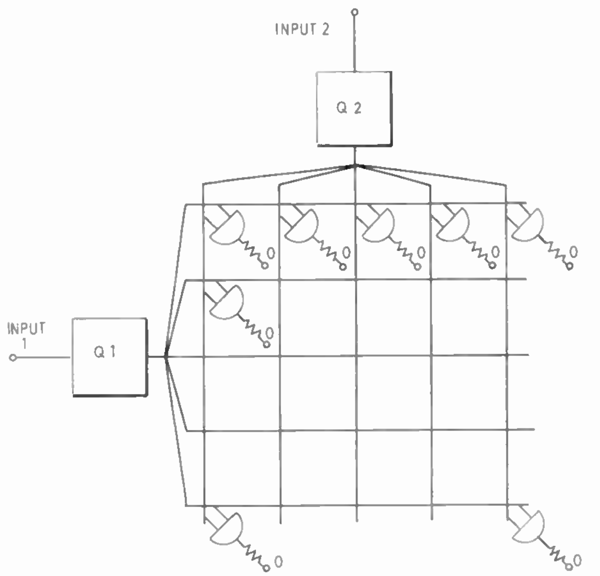


Fig. 9. Two-variable function generator with mutually exclusive setting using quantizers (O = summing junction).

face where n is the number of levels in the quantizer. We may also have $n \times m$ planes where the quantizers have m and n levels respectively.

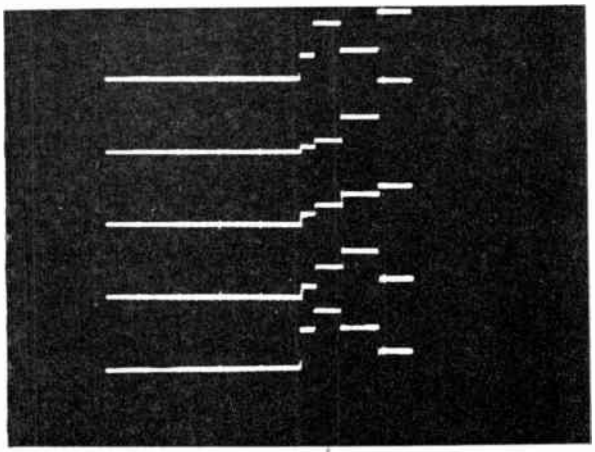
The quantizers necessary for this circuit may have very wide frequency responses since it is noted that fast 'kick sorters' used in analysis of energies of nuclear particles are of the same generic form. Thus it is to be expected that with a quantizer using a small amount of wide-band pre-amplification preceding a fast trigger circuit, the response of the quantizer would be measured in tenths of a microsecond. Thus provided the summing amplifier has wide bandwidth, there is a possibility that the frequency response of the function generator would be measured in terms of a few megacycles. The main points of difficulty in the circuit are in the large number of summing resistors, one for each point in the function, and the difficulties of adding currents from the large number of points. The only obvious solution to the problem is to sum partially in several computing amplifiers and then sum in a final stage.

6. Priority Sharing of a Computer between Several Processes

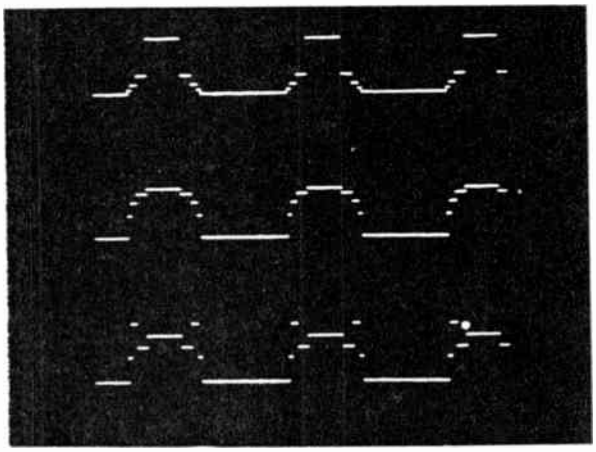
Time-sharing in data-processing is a well-established concept, and one of the motivations for its development was to avoid having a computer lying idle while comparatively slow input or output processes were in progress. Another objective was to allow a computer to carry on with less urgent work in the intervals between comparatively short programs on urgent problems. A quite different application of time-sharing has been in data-logging, where some form of multi-point switch scans consecutively a large number of information sources. (In some systems the sources are arranged in two groups, so that one group can be scanned more frequently than the other; but the whole process is repeated in a constant cycle.)

The time-sharing of a central computer between a number of processes or machines should preferably be on a priority basis, rather than by scanning in fixed sequences, the priorities being established by the current states of the devices which are being controlled. The channel or process which has maximum weighted error magnitude may be considered as that most urgently in need of control by the computer.[22]

The maximum channel is then selected, the computer samples the state variables of the process requiring correction, resets the integrator of this maximum channel, thus ensuring that another channel is selected in the next sample, computes the required correction, and applies it to the correct process to be



(*a*) Output vs input.



(*b*) Output vs time, sinusoidal input.

Fig. 10. Oscillograms from the quantized function generator.

controlled. This method ensures that no channel will be selected in successive samples unless the errors warrant it.

Thus a program (in the case of a general-purpose machine), or a piece of apparatus in the case of a specially-designed machine, must exist for extracting the maximum signal appearing at the input and, more important, must define in which channel the maximum amplitude occurs.

The selection may be accomplished by several methods. For example, a simple comparison-with-store circuit is obtained by slight modification of a quantizer of the type described in Section 3.

To recall the operation of the quantizer, we present to the input terminals the voltage to be quantized and if the input exceeds certain reference levels, a comparison amplifier or Schmitt trigger circuit turns ON. The trigger circuit corresponding to the highest quantum width causes all lower levels to switch OFF giving a set of mutually exclusive outputs. When the input voltage is reduced in magnitude each comparison circuit in turn switches OFF thus resetting the output to zero. If the Schmitt trigger comparison circuits are replaced by flip-flops, the flip-flop corresponding to each of the quantum levels exceeded will go ON and that flip-flop corresponding to the highest quantum level will turn OFF all lower levels, as before giving a set of mutually exclusive outputs. If the signal is removed then the highest flip-flop stays ON, thus remembering the signal magnitude.

If a new input is impressed on this quantizer which is in a state corresponding to the amplitude of the previous signal, then the state of the quantizer will be changed if, and only if, the state of a flip-flop corresponding to a higher quantum level is changed (i.e. input 2 > input 1). The quantizer thus selects the maximum quantum level irrespective of how many signals lie in this quantum width. A second sweep through the inputs will select those signals in this quantum width. The quantizer may then be reset to zero and the sequence started again. Thus we need two sweeps to select each maximum signal: the first selects the absolute magnitude and the second selects the channel in which it lies.

In a control application there is little physical significance in more than one signal lying in the maximum quantum width since if any one is selected (say the first reached in the second or identifying sweep) this will be removed from a number of subsequent sweeps.

A fast determination of the maximum signal can be made as follows. Let each signal be passed through either a quantizer of the type described in Section 3 or a de-coding tree, so that a different output line is energized for each amplitude of the signal. The system is then to employ a set of NOR gates such that each output line inhibits the lines for all signal channels corresponding to lower amplitudes. Consequently only the largest signal escapes being inhibited and is passed on.

The number of wires and the quantum widths need not be the same in each channel provided account is taken in the switching logic of the variation from signal to signal.

Modifications to the logic to admit more than one channel in the maximum quantum width could be made but since the hardware necessary in the circuit is liable to be large, this is regarded as uneconomic. For example, if we have m signals and n quantum widths in each, we have at least $m(m-1)(n-1)$ gates in the switching logic, omitting the quantizers and decoding trees.

This form of selection would be practical only in the case where not more than 5 processes were to be compensated with the highest possible speed, that is, for 32 quantum widths we would need approximately 600 gates. The selection could be accomplished in about 1 μs using conventional diodes and transistors and at a much higher speed using surface barrier, mesa or planar epitaxial transistors.

The attractiveness of the parallel computation lies in the high intrinsic speed of data processing.

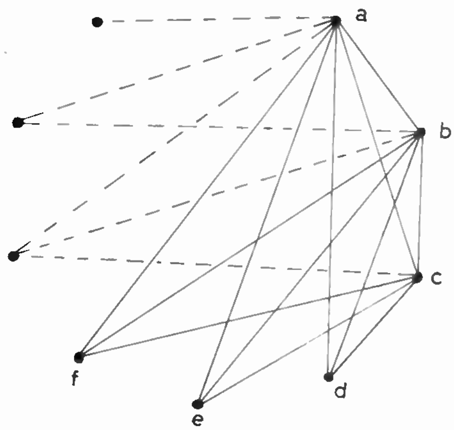
7. A Boolean Method for Selecting the Largest Member of a Set

A novel method of parallel selection is based on the properties of a certain Boolean matrix. If we are given a set of numbers which have to be ordered in decreasing order of magnitude then if they are labelled for convenience in alphabetic fashion a, b, c, d . . . and if, say, a is the maximum of the set, then axiomatically a is greater than b, a is greater than c, than d and so forth.

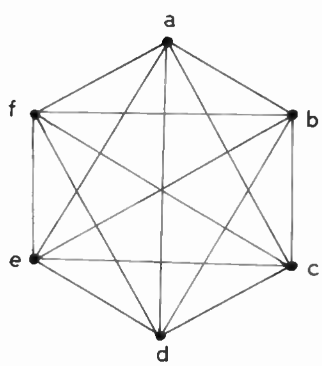
We define the ordered pair (a, b) to represent the Boolean vector result of comparing a with b in the sense that if a is greater than b, (a, b) is the logical 1, and if a is less than b, then (a, b) is the logical 0. Similarly, we have the opposite pair (b, a) which is 1 if b is greater than a and 0 if b is less than a.

Then if a is the maximum of the set, (a, b), (a, c), (a, d), (a, e) etc. must axiomatically all be 1. Thus the elementary product of the Boolean vectors (a, b), (a, c), (a, d) etc. will extract the maximum a.

This argument may be seen more clearly and be made more systematic with reference to Fig. 12 which shows the inputs a, b, c etc. arranged around the periphery of a circle. The ray connecting a to b represents the ordered pair (a, b). If we have n points distributed round the periphery, we will have $(n-1)$ rays emanating from any one point. If the Boolean



(a) Schematic Kuratowski graph of n variables.



(b) Complete 6-graph.

Fig. 11. Boolean matrix.

vectors represented by the rays emanating from any point are all 1 then that point must be the maximum of the set, and will be extracted by taking the elementary product of the Boolean vectors leading from that point.

The programme of such an extraction on a computer is facilitated by considering a matrix formulation of the Boolean vectors and for simplicity the concrete example of selecting the maximum of 6 inputs is taken. Starting with a, we may distribute the numbers clockwise round the circle and if a is maximum we require the vectors (a, b), (a, c), (a, d), (a, e) and (a, f) to be ON. This suggests a matrix formulation of the type

$$\begin{array}{ccccc} (a,b) & (a,c) & (a,d) & (a,e) & (a,f) \\ (b,c) & (b,d) & (b,e) & (b,f) & (b,a) \\ (c,d) & (c,e) & (c,f) & (c,a) & (c,b) \\ (d,e) & (d,f) & (d,a) & (d,b) & (d,c) \\ (e,f) & (e,a) & (e,b) & (e,c) & (e,d) \\ (f,a) & (f,b) & (f,c) & (f,d) & (f,e) \end{array} \quad \ldots\ldots(1)$$

Each row of the matrix corresponds to the rays leaving the node defined by the first letter in the ordered pair. Each column of matrix (1) has a unique graphical interpretation with respect to Fig. 11(b). Column 1 represents the vectors between first nearest neighbours and defines the periphery of the graph. Column 2 connects second nearest neighbours; column 3, third nearest neighbours, column 4, fourth, and so on.

If the matrix is written in its Boolean form with $(\overline{a, f})$ in place of (f, a), it will be found to be highly ordered and anti-symmetrical. If the elements of each row are used as inputs to a set of AND gates and if all the elements of one row are ON, that row will indicate which node is maximum.

If the second largest quantity is desired, we must remove the vectors involving the previous maximum node and as a consequence of the anti-symmetry of the matrix this may be achieved by removing the maximum line and the diagonal at 45 deg of positive slope starting immediately below the maximum line. Thus in the case where a is maximum this removal gives the matrix

$$\begin{array}{cccc} (b,c) & (b,d) & (b,e) & (b,f) \\ (c,d) & (c,e) & (c,f) & \overline{(b,c)} \\ (d,e) & (d,f) & \overline{(b,d)} & \overline{(c,d)} \\ (e,f) & \overline{(b,e)} & \overline{(c,e)} & \overline{(d,e)} \\ \overline{(b,f)} & \overline{(c,f)} & \overline{(d,f)} & \overline{(e,f)} \end{array}$$

This process is repeated as often as required.

If it is necessary for computational or practical reasons to cast out rows and columns, the matrix (1) must be rearranged and a subsidiary vector considered. It is noted that we are requiring that the diagonal to be cast out from matrix (1) must lie on the column of the new matrix. The matrix (1) involves only cross vectors, e.g. (a, b), (c, a), and to make the required matrix anti-symmetrical, which we must do to cast out rows and columns, we must define the self-products (a, a), (b, b), (c, c) etc. These ordered pairs have no absolute meaning according to the definition given previously, but they may be defined to be logical 1. Using this definition we may now write the square matrix equivalent of (1)

$$\begin{array}{cccccc} (a,a) & (a,b) & (a,c) & (a,d) & (a,e) & (a,f) \\ (b,a) & (b,b) & (b,c) & (b,d) & (b,e) & (b,f) \\ (c,a) & (c,b) & (c,c) & (c,d) & (c,e) & (c,f) \\ (d,a) & (d,b) & (d,c) & (d,d) & (d,e) & (d,f) \\ (e,a) & (e,b) & (e,c) & (e,d) & (e,e) & (e,f) \\ (f,a) & (f,b) & (f,c) & (f,d) & (f,e) & (f,f) \end{array} \quad \ldots\ldots(2)$$

The rows are now rays radiating *from* the nth node. The columns are now rays directed *towards* the nth node. The matrix is now anti-symmetrical about the main diagonal, i.e. in the usual matrix notation element a_{ij} = the negation of a_{ji}.

By examining the rows of this matrix and testing for all to be 1's the maximum signal will be extracted as before. The column corresponding to this row may now be removed or set to 1 and the process repeated to extract the maxima of successive subsets.

In practical circuits the pair (a, b) would be produced from analogue signals by an analogue computing amplifier of high gain having a differential input and a push-pull output clamped at voltage levels equivalent to 0 and 1. This arrangement may be used to set $a > b$ or $a \geqslant b$ to the logical level 1. The same circuit, because of the push-pull output, gives the negation signal (b, a).

In the case of digital inputs a full binary subtractor may be used as the comparator. In this case if a and b are the inputs and we subtract b from a, then if a carry comes from the most significant bit circuit b is the larger, otherwise a. Since merely the sign of the subtraction is desired, the circuit may be simplified by considering only the carry generator, starting from the most significant bit and terminating the computation whenever a difference occurs.

Once the $\sum(n-1)$ subtractions have been performed we may set up either matrix (1) or matrix (2) in a digital store.

Since each entry in the matrix is only a single binary digit, it may be possible to store a complete row in one computer word. (If the row is shorter than the computer word, any spare digit places should be filled with 1's. If it is longer, two or more words may be grouped together, as for double-length arithmetic.) To find whether a row contains all 1's, it is only necessary to collate the word containing that row with a comparison word containing all 1's, and test the result for numerical value zero.

After the correct row has been found it will also be necessary to eliminate the corresponding column. Setting all members of the column equal to one will have the same effect as striking out the column and closing up the matrix, and may be simpler. The procedure followed will depend on the logical facilities (e.g. end-around shift) available on the machine which is being used.

The advantages in a specially built machine are that (*a*) the comparison circuits could be simpler and possibly quicker than full binary subtraction, (*b*) the fast access memory needs only store bits and not words thus saving much storage space, (*c*) the maximum selecting circuit need only be one multi-input AND gate against the multiple gating of a general-purpose machine, and (*d*) the resetting of columns will be simpler.

8. Conclusions

The first objective in the application of on-line computer control is to put the feed-back information signals in digitally coded form so that they are both unambiguous and directly compatible with computer input and output. Therefore the first circuit developed was the Gray to binary code converter described in Section 2.

Position signals can usually be encoded by optical or magnetic devices, but many other variables (e.g. temperature) are usually represented by analogue electrical signals. The quantization of these signals is a step towards their digital representation. A quantizer was therefore developed, as described in Section 3.

The use of a high-speed computer on line can be economic only if its services are time shared between a number of machine tools or other controlled devices. The frequency of control action required by any one machine will vary from time to time according to the work in hand, so that it is not efficient to use a fixed cycle of scanning of all the controlled devices. Therefore some attention has been given to means of selecting the greatest member of a set of signals, so that priority may be given to it. Section 6 discusses methods based on amplitude inter-comparisons of the signals by analogue or threshold gating devices, while Section 7 describes a method which reduces all inter-comparisons to sign differences and then operates on a matrix formed of these sign differences.

As outlined in Section 1, the speed of the computer operations is vital, especially as one computer must be shared between a number of controlled variables in order to make the system economic. A number of the devices described are therefore designed with a view to fast working.

In the first place, the Gray to binary converter is parallel and asynchronous, so that it should be considerably faster than the more usual serial conversion of a signal sampled under clock control.

The next problem is the speed at which calculations relevant to an individual feed-back loop can be carried out. A common requirement will be to operate on some function of a variable (in simple cases the square or square-root); and speed may be increased by using analogue function generators. For a combination of speed and precision, quantized function generators have been developed. It is possible to make a quantized generator for functions of two variables, and there is the possibility that this function generator may be used in an adaptive control system as the reference model which gives the impulse response, step response or frequency response to which the overall dynamic system must adapt itself.

Alternatively, it is shown in the Appendix that calculations which look quite complicated can, with a little ingenuity in approximating and programming, be reduced to few enough computer orders to be evaluated in a time of the order of one millisecond.

If time-sharing is to keep the computer in use most of the time, there must be occasions when several controlled devices will require attention more or less simultaneously, and only one of them can receive it. To meet this situation, each controlled device must have a shut-down provision so that if it reaches the limit of tolerance before the computer accepts its signal, the device stops until the computer has taken corrective action. It should be noted that if the computer corrects channels in turn as fast as it can (instead of waiting for an error signal exceeding some threshold) the controlled variables will usually be kept very well within tolerance. On the occurrence of exceptional demands from several devices, all the others may drift somewhat further out in their tolerances (while the computer deals with the unusual group of priority demands) without reaching their shut-down limits.

In conclusion it may be said that all the elements needed to enable a central computer to exercise on-line control of a number of machine tools are available on paper. But the economics of such a system have not been analysed in detail, and its realization may have to await the occurrence of a problem which cannot be tackled by any other means.

9. References

1. N. Wiener, "Cybernetics", (Wiley, New York, 1948).
2. N. Wiener,"The Human Use of Human Beings" (Houghton, Mifflin, Boston 1950).
3. "Automation", (H.M. Stationery Office, London, 1956).
4. R. H. Macmillan, "Automation: Friend or Foe?", (Cambridge University Press, 1956).
5. At the Congrès International de l'Automatique, Paris, 1956.
6. M. V. Wilkes, "The second decade of computer development," *Computer J.*, **1**, p. 98, 1958.
7. Bourguet, "Optimalisation de la Conduite des Processus de Raffinage au moyen de Calculatrices Numériques Electroniques et d'Appareils d'Analyse en Continue: Etude de Rentabilité". Paper presented to a joint AICA, AFRA and AFCALTI colloquium, "Techniques Modernes de Calcul et Automatique Industrielle", Paris, 28th–31st May 1962.
8. A. L. Giusti, R. E. Otto and T. S. Williams, "Direct digital computer control", *Control Engineering*, June 1962.
9. I.B.M. Exhibit at the Interdata Exhibition, Munich, 1962.
10. Bernard and Deléglise, "Le Role et l'Utilisation des Calculatrices dans l'Automatisation des Unités Sidérurgiques." (See Ref. 7.)
11. V. M. Glushkov, in "Information Processing 62", C. M. Popplewell, ed., page 258 (North Holland Publishing Co., Amsterdam 1963).
12. "Conventional plant: automatic control", *Engineering*, **194**, p. 532, 26th October 1962.
13. "Förplanering för Jarnwerk" ("Planning for Steel Works"), *Facit Electronics Bulletin*, No. 2, 1962.
14. Csech, "Quelques Problèmes de Calcul Rencontrés dans l'Automatisation de Machines-Outils". (See Ref. 7.)
15. A. K. Susskind, ed. "Notes on Analogue-Digital Conversion Techniques", (Technology Press and John Wiley, New York, 1957).
16. B. Lippel, "Logical detenting in cathode ray coding tubes", *Trans. Inst. Radio Engrs (Instrumentation)*, I–7, No. 1, p. 29, March 1958.
17. R. W. Ketchledge, "Logic for a digital servo system", *Bell Syst. Techn. J.*, **38**, p. 1, 1959.
18. R. Wasserman and W. Nutting, "Solid state digital code-to-code converter", *Electronics*, **32**, No. 50, p. 60, 11th December 1959.
19. H. Huskey and G. A. Korn, "Computer Handbook", (McGraw-Hill, New York, 1962).
20. C. D. Morril and R. V. Baum, "Diode limiters simulate mechanical phenomena", *Electronics*, **25**, No. 11, p. 122, November 1952.
21. B. D. Smith, "Coding by feedback methods", *Proc. Inst. Radio Engrs*, **41**, p. 1053, August 1953.
22. P. M. Will, "Variable frequency sampling", *Trans Inst. Radio Engrs*, (*Automatic Control*) **AC–7**, No. 5, p. 126, October 1962.
23. A. T. Macdonald, "Torque- and velocity-limited servo systems", *Control*, **4**, p. 93, September 1961.

10. Appendix

The Time required for Digital Computations

As an example of the type of computation required, consider the problem of finding the switching point for an optimally-switched position control. This is a control in which maximum torque is applied in the forward direction up to a certain point and then maximum reverse torque, the point of reversal being so chosen that the load just comes to rest at the desired position. The equation relating angular travel θ after switching to velocity $\dot{\theta}_0$ at the moment of switching can be put in the form (Macdonald[23])

$$\theta = \frac{IT}{K_v^2}\left[\frac{K_v\dot{\theta}_0}{T} - \log_e\left(1+\frac{K_v\dot{\theta}_0}{T}\right)\right] \quad \ldots\ldots(3)$$

where I = inertia, K_v = coefficient of viscous damping, T = torque.

For digital computation this would be put in the form

$$\theta = L\dot{\theta}_0 - P\log_e(1+M\dot{\theta}_0)$$

where $L = I/K_v$, $M = K_v/T$ and $P = L/M$ are constants of the system. (Note that with this choice of coefficients there are no divisions involving the variable: division is the slowest of the four basic arithmetic operations on a digital computer.

Since $K_v\dot{\theta} = T$ at velocity saturation, we know that under all normal working conditions $M\dot{\theta}_0 < 1$. Hence the logarithm can be expanded as a series and

$$\theta = L\dot{\theta}_0 - PM\dot{\theta}_0 + \tfrac{1}{2}P(M\dot{\theta}_0)^2 - \tfrac{1}{3}P(M\dot{\theta}_0)^3 + \quad \ldots\ldots(4)$$

or approximately

$$\theta \simeq (L-PM)\dot{\theta}_0 + \tfrac{1}{2}PM^2\dot{\theta}_0{}^2 \qquad \ldots\ldots(5)$$

To handle optimal switching in the simplest way the computer would evaluate formula (5) at frequent intervals and reverse torque as soon as the predicted θ became greater than the difference between demanded position, θ_d, and present position θ_0. If the value of θ_0 is sampled at regular intervals the derivative $\dot{\theta}_0$ may be approximated by the difference between the present value θ_0 and the previous value θ_{-1}, the size of interval between samples being allowed for in the constants L and M in formulae (4) and (5). These constants L, M and P are functions of the equipment, varying slowly if at all with time, so that for each individual calculation the factors $(L-PM)$ and $\frac{1}{2}PM^2$ in formula (5) can be regarded as stored constants. The following tentative program is based on a single-address machine which for 12-bit words requires 120 µs maximum for multiplication and 20 µs for addition, subtraction and house-keeping operations. It is supposed that a new value of θ_0 is always read into storage location N_1; that the preceding value, denoted by θ_{-1}, is held in N_2; that the constant factors $L-PM$ and $\frac{1}{2}PM^2$ are held in N_{11} and N_{12}; and that the demanded position θ_d is held in N_{13}. Instructions are shown in symbolic form, since machine-code instructions would be relevant only to a particular machine, but it is assumed that each instruction operates on the contents of one accumulator X and one memory address N. A little time could be saved if one could do cumulative multiplications, and an expert programmer might find other savings when coding for a particular machine; but the total is approximately correct for a single-address machine of the assumed speed.

Instruction	Operation	Time in µs
1. $n_1 = \theta_0$	θ_0 to N_1	20
$x_1 = n_2$	θ_{-1} to X_2	20
$x_1 = x_1 - n_1$	Form $\theta_{-1} - \theta_0 = \dot{\theta}$	20
$n_3 = x_1$	$\dot{\theta}_0$ to N_3	20
$x_2 = n_1$ $n_2 = x_2$	θ_0 to X_2 and N_2	20
$x_p = n_3 \, . \, x_1$	Form $\dot{\theta}_0^2$ and	120
$n_4 = x_p$	store in N_4	20
$x_p = n_{11} \, . \, x_1$	Form $(L - PM)\dot{\theta}_0$	120
$n_5 = x_p$	and store in N_5	20
$x_1 = n_4$	Bring out $\dot{\theta}_0^2$	20
$x_p = n_{12} \, . \, x_1$	Form $\frac{1}{2}PM^2 \, \dot{\theta}_0^2$	120
$n_4 = x_p$	and store in N_4	20
$x_1 = n_4$	Add $(L - PM) \, \dot{\theta}_0$	20
$x_1 = x_1 + n_5$	and $\frac{1}{2}PM^2 \, \dot{\theta}^2$	20
$n_6 = x_1$	Store θ in N_6	20
$x_2 = n_{13} - x_2$	Form $\theta_d - \theta_0$	20
$x_2 = x_2 - n_6$	Form $(\theta_d - \theta_0) - \theta$	20
Jump to 20 if $x_2 < 0$	and test if < 0	20
Go to instruction 1	Repeat sequence	(20)
20. Output	Operate switch	(20)
		700

(After the jump test, only one of the two instructions with time in brackets will be executed.) The next term in the approximation would be $-\frac{1}{3}PM^3\dot{\theta}_0^3$, and this could be included without bringing the total time up to 1000 µs.

The time could be drastically cut by using analogue multiplications for the derivative $\dot{\theta}_0$ and retaining the rigid precision of digital working only for the positional information. $\dot{\theta}_0$ could then be obtained from a tachometer generator and scaled and squared in analogue circuits. Formula (5) would thus be evaluated by analogue methods to the desired number of terms, and the resulting value of θ digitized and compared with θ_d and θ_0 as in the last five instructions in the program above. The computing time is then cut down to 80 µs plus the encoding time.

Manuscript first received by the Institution on 3rd January 1963 and in final form on 26th March 1963 (Paper No. 833/C59)

DISCUSSION

Under the chairmanship of Professor A. D. Booth

Mr. S. L. H. Clarke: Dr. Will referred to priority interrupt facilities on a number of computers, saying that he felt that such systems where priority levels are set once for all were not necessarily applicable for industrial process control. Several computers that he did not mention but which are well known to me, have a very simple interrupt system from the hardware point of view which, through the medium of a control word, allows the program to allocate and then observe, different priority levels. In many applications the priority of different interruption systems vary with the time of day, stage of process batch or phase of a mission. This change of priority level could, of course, be achieved by hardware, but the flexibility of program control will be the most economical way of achieving the required result.

The authors (*in reply*): The examples of specific types of computer mentioned during the presentation of the paper were for illustration only and not meant to be exhaustive.

We agree with Mr. Clarke that program control gives great flexibility and would point out that in several places in the text we have been careful to show that the whole operation of priority sharing of the form enunciated here could be performed on a general purpose machine. The design of the hardware or the writing of the program depends upon the same basic principles (e.g. Kuratowski graph formalism (Section 7)) and the object of this part of the paper was to describe the principles.

In our view the method of allocating a control word to describe priority would become that described here if the control word were changed *at the end of each iteration by the computer*—thus the control word technique as used at present may be looked upon as the quasi-steady state mode of the more complex dynamic priority allocation considered in Section 6.

Colour Television Systems on Trial

A series of demonstrations of colour television systems were given by the British Broadcasting Corporation and other broadcasting organizations and manufacturers during July to members of the European Broadcasting Union, representatives of the O.I.R.T. (the corresponding organization in Eastern Europe) and representatives of the C.C.I.R. (The International Radio Consultative Committee). When all the results of these and other appraisals, together with the results of the field trials, have been studied, it is hoped that the Western European broadcasting authorities will be able to arrive at a definite recommendation on the choice of system at an E.B.U. meeting to be held in October. This recommendation will be considered, and it is hoped accepted, by the Telegraph and Telephone Administrations of the whole of Europe at a special meeting which may be convened by the C.C.I.R. in February 1964.

B.B.C. Demonstrations

The selection from the main series of demonstrations given recently to the Technical Press by the B.B.C. could obviously do no more than show a limited number of aspects of a very complex problem. Colour pictures and also black-and-white compatible pictures were shown using the three systems: N.T.S.C., Secam and P.A.L. It is important to compare the picture quality not only under ideal conditions but also in the presence of the various kinds of distortion and interference that may occur when the pictures are received in the home. Such disturbances may originate in the studio equipment, in the process of video-tape recording, in transmission over long distribution networks, in the radio transmitter, in the course of radio propagation and in the receiver itself.

The demonstrations illustrated the effect of some typical types of disturbance on the pictures obtained on each of the three systems (the pictures were all 625 lines). The effects of ignition type interference, signal strength reduction, bandwidth limitation and video recording were shown for the three systems: none seemed more than marginally better or worse than the others. The effects of transmission over long lines and the ensuing phase delays were not demonstrated.

It was not possible to draw final conclusions on the overall merits of the three systems from such a short series of demonstrations. To do so, it would be necessary to take a great many factors into account, including the conditions that are likely to be obtained in actual practice, which may affect relative service area and—most important—the cost, reliability and ease of adjustment of colour receivers designed for each of the three systems. When colour pictures are first introduced in this country only a minority of viewers will be equipped with colour receivers; the quality of the compatible pictures received on black-and-white receivers is therefore an important consideration.

To make a final judgement and a decision on the best system, it is necessary to make extensive field trials under all likely conditions of transmission and reception. Such a series of trials is being made in this country in co-operation between the B.B.C., the Post Office, the I.T.A./I.T.C.A. and the Radio Industry. Since last September the B.B.C. has transmitted N.T.S.C. colour pictures from the u.h.f. transmitters at the Crystal Palace station on 625 lines. Since April 1963 the Secam system, developed by the Compagnie Française de Television has also been used in these trials; transmissions on the P.A.L. system, developed by the Telefunken Company in West Germany, were introduced into the trials in June. As two u.h.f. transmitters are available at Crystal Palace station, operating on different channels, it has been possible to make direct comparisons between the N.T.S.C. system and both of the other systems.

Trials and demonstrations have also taken place in some other European countries and the B.B.C. has taken the initiative in encouraging the co-ordination of this work among members of the European Broadcasting Union in order to pave the way for an early decision on a common system of colour television to be used throughout Europe. It is important that all European countries should use the same system, to make it possible for live exchanges of programme in colour to take place. As far as Trans-Atlantic exchanges are concerned there are inherent difficulties in conversion from 525 to 625 lines and vice versa and although changes to a different *system* would add to the difficulties, the increased complexity is understood to be not excessive. Time differences on either side of the Atlantic Ocean may well prove greater deterrents to direct programme exchanges than the technical problems.

The B.B.C. states that if a decision is reached at the C.C.I.R. meeting in February next year in favour of the N.T.S.C. system, it would be possible to introduce a limited amount of colour television into B.B.C. 2 early in 1965. If either of the other systems is chosen it will take somewhat longer to equip the studios and for the Industry to produce receivers on a large scale, probably until early 1966.

Separate Luminance Colour

The demonstration of colour television equipment given in London by Marconi's Wireless Telegraph Company and English Electric Valve Company, showed four types of colour television camera in operation at the demonstration. Two of these cameras were operating on the present-day standard system of three camera tubes which produce red, green and blue signals, corresponding to the three primary components of the colour picture. The other two demonstrated 'separate luminance' signals. These equipments were all shown to be capable of giving good quality television pictures with any of the three coding systems at present being evaluated by delegates from the European Broadcasting Union in an attempt to agree upon a common European standard for colour television.

In the first of these new 'separate luminance' cameras, four separate tubes are used. One produces a high quality luminance signal while the other three produce the red, green and blue signals which are then combined to give the two colour, or chrominance, signals used in transmission. This enables the luminance signal to be used in black and white receivers to give a much higher quality picture.

The second 'separate luminance' camera operates on similar principles but has only three camera tubes. This development reduced the size and complexity of the camera channel and was stated to bring all the advantages of separate luminance systems. The three tubes used are all 3-inch image orthicons, and the red and blue components of the light input are directly converted into electrical signals in two of these tubes while the luminance signal is produced separately by the third tube. The green component can then be derived from these three signals. Any of the standard coding systems may then be used to produce the transmitted signals. The scene illumination required for this type of camera is only one third of that required for existing colour cameras, and compares with black and white studio requirements.

Similar demonstrations of a separate luminance system were given to the E.B.U. delegates by Electric and Musical Industries but here the four-tube approach was firmly advocated. A 4½-in. image orthicon was used for the luminance signal with three 1-in. vidicons to derive the chrominance signals. A more satisfactory solution to the compatibility problem and a cheaper camera were claimed to be the outstanding advantages of this technique over the three-tube camera. Separate luminance would certainly seem to offer considerable advantages over the conventional approach when the scene lighting requirements are considered. From the receiver point of view, with the four-tube camera it is possible to obtain *acceptable* pictures without any modification, the slight change of saturation in red areas being appreciable only when direct comparison can be made. Ideally, however, a multilinear matrix encoder is required which could be made relatively cheaply.

NOTE ON COLOUR SYSTEMS

The N.T.S.C. system, evolved by the National Television System Committee in the United States in 1953, is now familiar to most British radio engineers. (A description was given in the paper "Colour television" by G. N. Patchett in the *Journal* for November 1956.) The Secam and P.A.L. systems are not so well known and the following brief details are taken from *B.B.C. Record* for May 1963.

The Secam System

The Secam system was developed some years ago in France, by the Compagnie Française de Television, its name being a combination of 'sequential' and 'memory', and it has recently been considerably improved. The red, blue, and green components are generated in the same way as in the N.T.S.C. system; the 'luminance' signal and the two colour signals are also derived in the same way. The two colour signals are, however, transmitted not simultaneously but consecutively during alternate line periods. This enables the Secam system to give satisfactory colour rendering in the presence of larger inequalities in the transmission of the two colour signals than can be tolerated by the N.T.S.C. system. This is an advantage when the pictures have to be transmitted over extensive distribution networks or recorded on magnetic tape. On the other hand, the picture definition given by the Secam system is theoretically less good than that of N.T.S.C., and Secam may prove somewhat more susceptible to interference caused by excessive noise—appearing as a moving granular background to the picture. The Secam system requires a delay line in the receiver to store the colour signal transmitted during one line period so that it becomes available during the next line period simultaneously with the colour signal transmitted during that line period. The Secam system also differs from the N.T.S.C. system in that the sub-carrier is frequency-modulated by the colour signal, instead of amplitude-modulated. As in the N.T.S.C. system the receiver derives the red, blue, and green components from the colour signals and the 'luminance' signal and applies them to a shadow-mask tube.

The P.A.L. System

The P.A.L. system has recently been developed by the Telefunken Company in West Germany. It is a variant of the N.T.S.C. system and is also, like the Secam system, intended to reduce the susceptibility of the colour rendering to inequalities in the transmission of the two colour signals. As in the N.T.S.C. system these two signals are transmitted simultaneously by amplitude modulation of a sub-carrier, but one of them is reversed between alternate lines; hence the name P.A.L. (phase alternation line). By this means errors in colour, such as those mentioned above, are averaged out between one line and the next and this averaging process may be assisted by using a 'delay line' in the receiver, as in the Secam system. Apart from this, and the switching arrangements for reversing one of the colour signals, the receiver is basically the same as an N.T.S.C. receiver.

U.D.C. 621.317.76

A Frequency Meter with Continuous Digital Presentation

By

P. WOOD†

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

Summary: Many transducers produce an alternating output with a frequency proportional to the transduced quantity. The normal method of converting this output into a suitable digital form is by counting the number of cycles that occur during a fixed time. This paper describes an alternative approach in which the binary number corresponding to the input frequency is continuously available.

The system is based on the use of a reversible binary counter. The input frequency is converted into a suitable train of pulses and applied to the ADD input of the counter, and a second train of pulses whose mean frequency is proportional to the number stored into the counter to the SUBTRACT input. When equilibrium is established the number stored in the counter is proportional to the input frequency.

1. Introduction

Some types of transducer used in measurement and control applications generate an electrical signal with a frequency proportional to the magnitude of the quantity being measured. One example of this is the vibrating cylinder pressure transducer, in which a thin-walled metal cylinder forms a mechanical resonant system of very high Q, the natural frequency changing with applied pressure. A closed-loop system causes the cylinder to vibrate at the natural frequency, the frequency giving a measure of the applied pressure. For use with digital control systems, and for indicating purposes, the frequency must be converted into an equivalent binary or decimal digital representation.

The conventional method of achieving this is to count the number of cycles occurring during an accurately determined time interval; the contents of the counter at the end of the period give the digital representation of the frequency. This method suffers from the disadvantages that a true indication of frequency is not available during the counting period, and that a change in frequency does not become apparent until the next counting cycle has occurred. Also in some high-speed sampled data control systems it is not possible to obtain the required degree of accuracy as an insufficient number of cycles occurs during the sampling period.

To overcome these disadvantages a simple experimental system has been produced which gives a continuous digital indication of frequency. This gives a six-binary digit display of any frequency in the range 250 c/s–16 kc/s, but other frequency ranges and a higher degree of accuracy can be easily obtained, while the circuits can be modified to give a decimal read-out.

† The Plessey Co. (U.K.) Ltd., Roke Manor Research Laboratories, Romsey, Hampshire.

2. Basic Principles

The basis of the continuous display frequency indicator is a reversible binary counter. In such a counter, gates between each stage ensure that a 'carry' is generated when the preceding stage changes from '1' to '0' for an ADD input, and from '0' to '1' for a SUBTRACT input. The variable frequency to be measured is converted into a train of pulses and applied to the ADD input.

The SUBTRACT input is obtained from a binary rate multiplier‡ comprising a binary scaler to which is applied a fixed frequency pulse train, the outputs of the scaler being fed to a series of gates controlled by the stages of the binary counter. The outputs of these gates are combined and give a train of pulses with a mean rate equal to the product of the fixed frequency and the contents of the counter (scaled as a fraction). When the counter reaches equilibrium, the mean pulse rates applied to the ADD and SUBTRACT inputs must be equal, and hence the number held by the counter gives the digital representation of the input frequency. If this frequency changes an excess of ADD or SUBTRACT pulses results, leading to the number in the counter changing until equilibrium is re-established.

A number of problems have to be overcome to ensure a stable indication of frequency. Firstly, if an

‡ E. M. Grabbe, S. Ramo, D. E. Wooldridge, "Handbook of Automation, Computation and Control", Vol. 2, p. 29–35 (John Wiley, New York, 1959).

ADD and a SUBTRACT pulse overlap, the effect on the counter will be indeterminate, and the indicated count will be subject to jitter. A coincidence detector is therefore necessary to inhibit the two pulses when they coincide. Secondly, in the equilibrium state ADD and SUBTRACT pulses would alternate, causing the indicated count to alternately increase and decrease by one digit. To overcome this circuits should be included to detect alternate pulses of opposite sign, and ensure that they are not fed to the counter. Thirdly, the counter is only able to indicate discrete frequencies, whereas the actual frequency could have any values within the designed range. Intermediate frequencies result in the counter alternating between the discrete values on either side of the actual frequency due to a random phase relationship between the unknown input frequency and the internal standard. This can be overcome by phase-locking the two pulse trains once each complete cycle of the binary scaler.

3. Method of Operation

A block diagram of the complete system is shown in Fig. 1, and typical waveforms appearing at various parts of the circuit in Fig. 2.

The reversible binary counter contains six stages with an ADD/SUBTRACT carry gate between each stage. The output of a stable 16 kc/s clock pulse generator is applied to a binary scaler, each stage producing a pulse train at half the frequency of the preceding stage. These pulse trains are applied to gates, each of which is controlled by the corresponding stage of the reversible counter. The output of these gates, combined by an OR gate, triggers a multivibrator (the SUBTRACT pulse shaper) to produce a train of positive and negative 30 μs pulses. In Fig. 2, the clock pulses are shown on line C and the inputs of each stage of the binary scaler on lines D–J. A typical combined pulse train appears on line K, and the outputs of the subtract pulse shaper on lines L and M.

The output of the transducer is applied to the ADD pulse shaper, which converts the incoming sine wave into anti-phase trains of 30 μs pulses (A and **B** in Fig. 2). If pulses from the ADD and SUBTRACT pulse generators coincide the anti-coincidence gate generator is energized and produces a 50 μs pulse. The trailing edges of the ADD and SUBTRACT pulse generator outputs trigger 2 μs multivibrators, these 2 μs pulses being combined in an OR gate. If a pulse is produced by the anti-coincidence gate generator the output of the OR gate is inhibited, otherwise it triggers a third 2 μs multivibrator. An output pulse is therefore only produced when the ADD and SUBTRACT pulses do not overlap.

The trailing edge of the negative going output from the subtract pulse shaper also triggers a multivibrator to produce a 10 μs pulse which operates the ADD/SUBTRACT gates in the reversible counter. In Fig. 2, the anti-coincident generator output is shown on line N, the input of the third 2 μs multivibrator on line P and the ADD/SUBTRACT gate input on line Q.

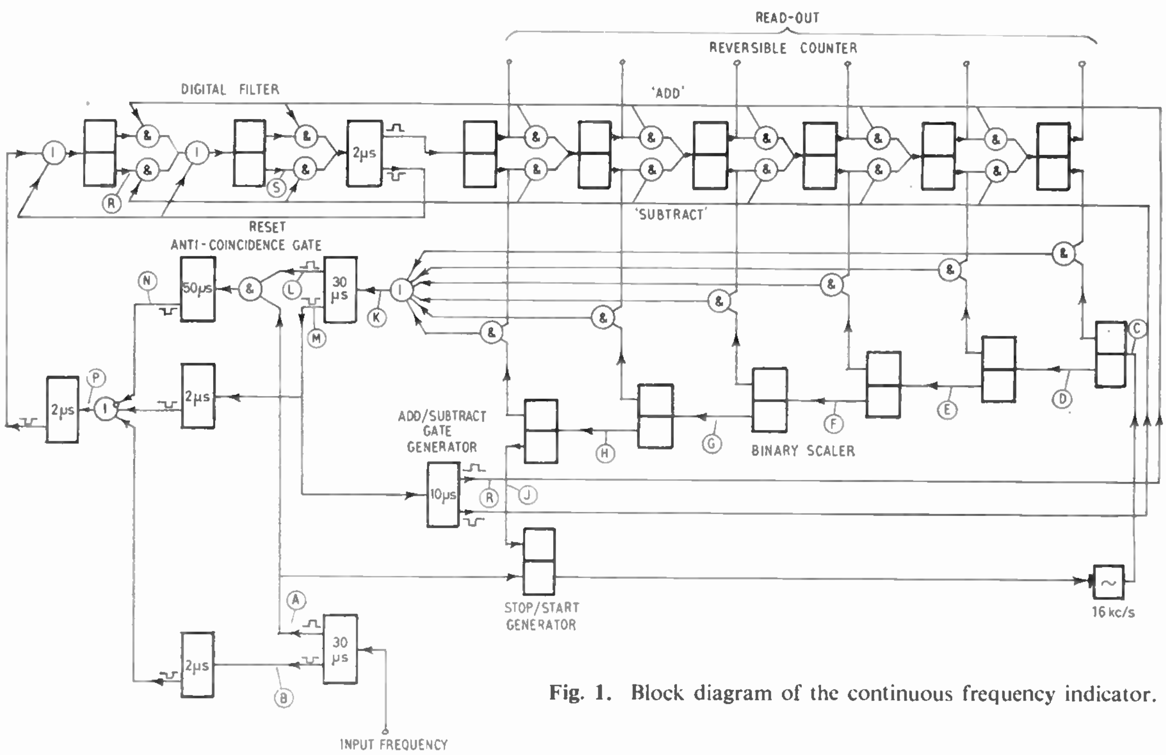


Fig. 1. Block diagram of the continuous frequency indicator.

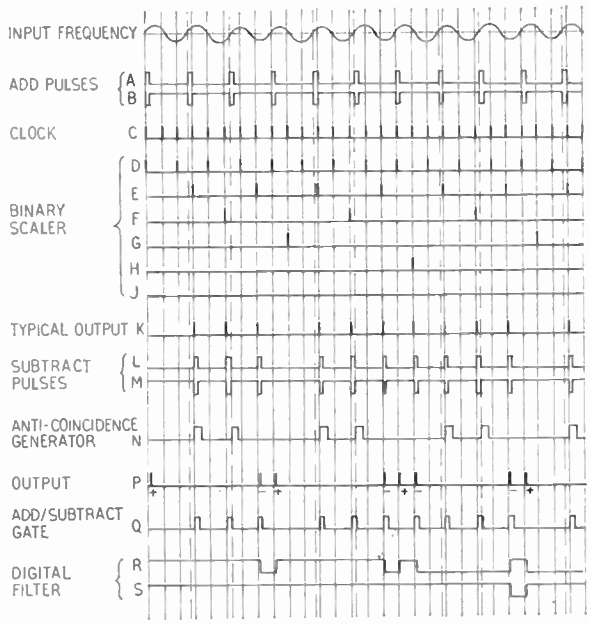


Fig. 2. Waveforms of the continuous frequency indicator.

The final part of the circuit, which is referred to as the digital filter eliminates jitter of the counter due to the generation of alternate ADD or SUBTRACT pulses. The filter requires an excess of two pulses of the same polarity to produce an output, and hence after equilibrium has been reached alternate pairs of ADD and SUBTRACT pulses are also eliminated.

Table 1

Input	Filter	Output	Remarks
+	11		
−	10		
+	11		
−	10		
+	11		
+	00	+	Equilibrium reached
	11		Digital filter set to 11 by + output
−	10		
+	11		
−	10		
+	11		
−	10		Pair of negative pulses suppressed
−	01		
+	10		
−	01		
+	10		
−	01		
+	10		Pair of positive pulses suppressed
+	11		
−	10		
+	11		

The digital filter comprises a two-stage reversible binary counter, the input to this coming from the anti-coincidence circuit. When a 'carry' is generated by the second stage (resulting either from an ADD or a SUBTRACT), an output is produced which is applied to the first stage of the main reversible counter. The two stages of the digital filter are re-set to a condition which depends upon whether the carry was generated as the result of an ADD or SUBTRACT pulse. Table 1 shows an example of how the digital filter reaches equilibrium, and how subsequent alternate ADD and SUBTRACT pulses are inhibited. It is assumed that initially the main counter is one digit below its stable value.

If a 'carry' has been generated as a result of the digital filter changing from 00 to 11 on the receipt of a SUBTRACT pulse, it is reset to 00. The change of state of two stages of the digital filter is shown on lines R and S of Fig. 2; since the counter is shown in a state of equilibrium no output is produced during the period shown.

4. Basic Circuit Elements

No attempt is made in this paper to give full circuit details as only experimental circuits were produced, but the basic elements which formed the system will be described.

4.1. *Binary Counter*

The basic circuit of a binary counter stage with its associated ADD/SUBTRACT gate is shown in Fig. 3. Each binary stage is a conventional transistor bi-stable multivibrator, the digit levels being established at approximately 0 V and −4 V by the bottoming of the transistors and the catch diode connected to the −4 V rail. The counter is defined as containing a '1' when the right-hand side of the bistable is at 0 V, and correspondingly registers a '0' when it is at −4 V.

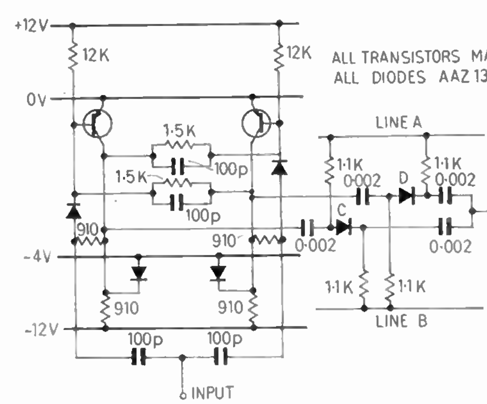
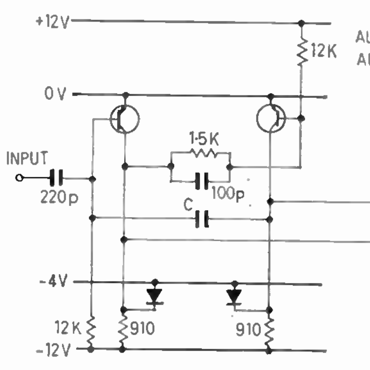


Fig. 3. Counter stage and ADD/SUBTRACT gate.

The ADD/SUBTRACT gate is controlled by the output of the ADD/SUBTRACT gate generator through lines A and B. When an ADD pulse is applied to the first stage, line A is held positive with respect to B and diode C conducts. If the input to the bistable stage causes the left-hand side to rise from -4 V to 0 V, the voltage step is differentiated and fed through the diode to produce an input to the following stage. On the other hand if a SUBTRACT pulse is applied to the input, the polarities of A and B are reversed, diode D conducts, and a positive pulse is transferred to the next stage if the right hand side of the bistable multivibrator switches from -4 V to 0 V.

4.2. *Monostable Flip-flop, Pulse Shaper and Delay*

The monostable flip-flop, pulse shaper, and delay element is shown in Fig. 4. This circuit closely resembles the bistable flip-flop, one of the capacitor/resistor couplings between collector and the opposite base being replaced by a capacitor with a value dependent upon the pulse width required. Positive-going pulses are used to trigger the multivibrator; a delayed trigger is available by differentiating the positive-going trailing edge of the pulse produced by the normally 'on' collector of the monostable multivibrator.



PULSE WIDTH	C
2µs	470p
10µs	1800p
30µs	5600p
50µs	0·01µ

Fig. 4. Monostable flip-flop.

4.3. *Anti-Coincidence Gate*

The anti-coincidence gate and associated circuits are shown in Fig. 5, together with some waveforms to clarify the operation. The inputs from the transducer and binary scaler are applied to the two pulse-shaping monostable multivibrators to produce 30 µs pulses. The positive-going outputs A and B are combined in a 'diode-resistor' AND gate; if positive outputs from both multivibrators are present at the same time point A goes positive, triggering the inhibit gate generator multivibrator producing a pulse of 50 µs width (C).

Two further multivibrators are triggered to produce 2 µs pulses (D and E) at the conclusion of the 30 µs pulses from the ADD and SUBTRACT pulse shapers, and either of these pulses would normally trigger the third multivibrator. If, however, a pulse from the inhibit gate generator is present, it results in points H and J becoming more negative, and a positive pulse from multivibrators D and E applied to point J is of insufficient amplitude to trigger the multivibrator. The output pulse is therefore inhibited.

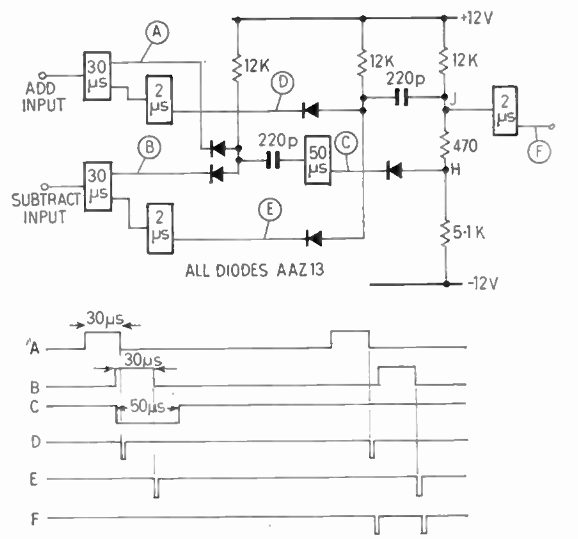


Fig. 5. Anti-coincidence gate—circuit and waveforms.

5. Binary and Decimal Indication

The experimental system described uses a subminiature voltage indicator type DM160 to indicate the state of the stages in the reversible counter. These have the advantage that the grid swing from cut-off to conduction is directly compatible with the voltage levels at the collectors of the transistor stages, and the input impedance is high enough to prevent loading effects. These indicators are adequate if a binary indication is required under conditions of normal ambient illumination.

The simplest method of providing a decimal indication is to provide a separate reversible counter connected to count in a scale of 10 and supply it with the same input as the reversible binary counter which is used to control the frequency divider.† Although this leads to an increase in circuit complexity, it is often convenient to have both a decimal number for display purposes and its binary equivalent for application to a computer.

6. Comparison with Existing Techniques

Compared with existing techniques the method described in this paper would apparently require a larger number of components without offering any compensating advantages for normal use. However a conventional counter would normally have to use a

† J. R. Goldberg, "Reversible decimal counters", *Electronic Technology*, 38, pp. 234–45, July 1961.

divider counting down from a stable high frequency oscillator to establish the accurate time interval required, and therefore the actual number of components used would not differ significantly between the two methods.

For some purposes, the fact that a change in input frequency appears without a significant time lag may be advantageous; on the other hand if the input frequency is subject to jitter the resultant changes in the indication may lead to difficulty in establishing the mean value.

The main application of the method described in this paper is in sampled data digital control systems with high sampling rates and short sampling periods. For transducers with relatively low frequency outputs (say below 100 kc/s) the frequency cannot be measured to the required degree of accuracy because there is insufficient time to allow an adequate number of cycles to accumulate in the counter. One conventional method of overcoming this is to count over a relatively long time interval and to transfer the number held in the counter to a buffer store at the conclusion of the counting sequence. It is then available for use in the control system until the next counting sequence has been completed. This approach is less economical in components than that using a reversible counter, and also suffers from the disadvantage of delay between a change in the transducer output frequency and its digital representation being available.

7. Response Time

If the reversible counter initially contains all zeros and an unknown frequency is then applied, there will be a time delay before the counter stabilizes at its final value. It is apparent that the rate at which pulses are applied to the ADD input is proportional to the difference between the unknown frequency and the stored count, and hence the rise in the indicated frequency approximates to an exponential until the final value is reached. The time-constant of this rise is equal to the ratio between the maximum capacity of the counter divided by the frequency of the fixed frequency pulse train applied to the binary scaler.

In practice the input frequency is unlikely to change instantaneously (due to the finite bandwidth of the transducer and the dynamics of the system to which the transducer is applied), and the delay will generally be unimportant.

8. Conclusions

To obtain an accurate measurement of frequency by counting the number of cycles occurring in a given time requires a relatively long time interval. This is a disadvantage in sampled data control systems where sampling times are short, and it results in a separate buffer store being necessary to hold the previous contents of the counter while a new counting sequence is proceeding. To overcome this a system has been described in which a binary number proportional to an input frequency is continuously available to any required degree of accuracy.

This approach is particularly useful for process control using transducers which generate a frequency proportional to the magnitude of the controlling parameters. The outputs of such transducers are available for sampling at any time, a continuous binary or decimal indication is available for monitoring purposes, and the indicated frequency follows changes in the transduced parameter with a negligible delay.

The principles of the system have been proved by the construction of an experimental six-stage binary indicator to measure frequencies up to 16 kc/s and it is intended to incorporate a developed version in the electronic section of a novel mass flowmeter.

9. Acknowledgments

The author wishes to thank Mr. D. Underwood, who developed the circuits used in the application, and the Directors of the Plessey Co. (U.K.) Limited, for their permission to publish the paper.

Manuscript received by the Institution on 10th December 1962. (Paper No. 834.)

POINTS FROM THE DISCUSSION

Mr. A. A. Fayers: I would suggest that it is not possible to extend the number of bits (actually 500 τ units) indefinitely over the range stated—perhaps the extension of 10 bits at the most could be achieved. Secondly, in my view, the processing system is incapable of realizing the full accuracy of the transducer.

The Author (*in reply*): When information is converted from an analogue (or continuous) to a digital (or incremental) form, analogue inputs that occur between the discrete levels established by the digital process must have the digital representation corresponding to one of the discrete levels bounding the analogue input. In the case of frequency, an asynchronous input would result in the indication changing randomly between these levels; the system described in the paper eliminates the unstable indication and represents intermediate frequencies by the quantized frequency immediately below the unknown.

The maximum error that can occur is slightly less than one quantum; this error can be reduced by increasing the number of bits and theoretically this can be carried on indefinitely. However 10 bits would appear to be the optimum since the maximum error (just under 0·1% of full scale) is compatible with the accuracy of the transducers with which this system is intended to operate.

INSTITUTION NOTICES

Institution Premiums and Awards

The Council of the Institution announces that the following awards are to be made for outstanding papers published in the *Journal* during 1962:

CLERK MAXWELL PREMIUM

"Propagation Influences in Microwave Link Operation" by M. W. Gough, M.A. (July *Journal*).

ASSOCIATED REDIFFUSION PREMIUM

"A Constant Luminance Colour Television System" by I. J. P. James, B.Sc. and W. A. Karwowski, B.A. (April *Journal*).

V. K. ZWORYKIN PREMIUM

"Auditory Perception and its Relation to Ultrasonic Blind Guidance Aids" by L. Kay, B.Sc., Ph.D. (October *Journal*).

A. F. BULGIN PREMIUM

"Electron Transit Time and other Effects in a Valve Voltmeter Operating at Extremely Low Current" by I. A. Harris (November *Journal*).

LESLIE MCMICHAEL PREMIUM

"Optimum System Engineering for Satellite Communication Links with special reference to the Choice of Modulation Method" by W. L. Wright, B.A. and S. A. W. Jolliffe (May *Journal*).

CHARLES BABBAGE AWARD

"Tunnel Devices as Switching Elements" by I. Aleksander, B.Sc.(Eng.), and R. W. Scarr, Ph.D., B.Sc.(Eng.), (March *Journal*).

LORD RUTHERFORD AWARD

"Semiconductor Nuclear Radiation Detectors" by G. Dearnaley, Ph.D. (August *Journal*).

MARCONI PREMIUM

"Switchable 405/625 Line Time-Bases" by A. Ciuciura, B.Sc.(Eng.) (October *Journal*).

ARTHUR GAY PREMIUM

"Automatic Component Assembly in the Telephone Industry" by D. Hinchcliffe, J. R. W. Smith, M.Sc. and G. H. King, B.Sc. (September *Journal*).

DR. NORMAN PARTRIDGE MEMORIAL PREMIUM

"A Pulse Time Multiplex System for Stereophonic Broadcasting" by G. D. Browne (February *Journal*).

J. C. BOSE PREMIUM

"A Two-State Device with Two Inductively Coupled Colpitts Oscillators" by B. R. Nag, M.Sc.(Tech.), M.S., D.Phil. (July *Journal*).

The following Premiums are withheld because either no eligible papers were published during the year or the eligible papers were not of high enough standard: The Heinrich Hertz Premium, the J. Langham Thompson Premium, the Lord Brabazon Award, and the Hugh Brennan Premium.

The Premiums and Awards will be presented at the Annual General Meeting in London on 27th November 1963.

International Telemetering Conference

The first International Telemetering Conference, sponsored jointly by the American Institute of Aeronautics and Astronautics, and American Institute of Electrical and Electronics Engineers, the Instrument Society of America, the British Institution of Radio Engineers and the Institution of Electrical Engineers, will take place in London from the 23rd to 27th September 1963, at the I.E.E. headquarters in Savoy Place.

Delegates will be able to see a fully representative selection of equipment at the International Telemetering Exhibition, which will run concurrently with the Conference and be held at the London Hilton Hotel. Tickets for the exhibition will be issued to all who register for the Conference.

The Conference subjects will be divided into the following sessions:

Monday	3.30 p.m.	Introductory
Tuesday	9.30 a.m.	Industrial systems
	2.15 p.m.	Transducers and signal conditioning
	6.00 p.m.	Geophysical and biomedical systems
Wednesday	9.00 a.m.	Recording and data processing
Thursday	9.30 a.m.	Modulation, coding and multiplexing
	2.30 p.m.	Error detection demodulation and synchronizing (including random sequences)
Friday	9.30 a.m.	Aerospace systems

Fuller details can be obtained from the Secretary of the Institution, 9 Bedford Square, London, W.C.1.

Copies of 1963 Convention Papers

Most of the papers read at the Convention are being considered by the Papers Committee for publication in *The Radio and Electronic Engineer* and will appear in issues during the remainder of 1963.

Complete sets of preprints of the papers, as prepared for delegates at the Convention, are still available, price £3 15s. for the complete set. The papers are arranged in three separate Parts according to Session and may be purchased at 25s. each Part:

PART A: Survey papers on economic and theoretical aspects. Sensing Devices, Measurement and Telemetry.

PART B: Control and Information Processing.

PART C: Industrial Applications of Electronic Systems.

Certain individual papers may also be obtained from the Institution's Publications Department, price 2s. 6d. per copy. A full list of the papers including summaries was given in the February and March issues of *The Radio and Electronic Engineer*.

U.D.C. 621.396.677

The Angular Resolution of a Receiving Aperture in the Absence of Noise

By

V. G. WELSBY, Ph.D.†

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

Summary: The paper is concerned with a more general approach to the problem of the performance of particular directional receiving arrays and their associated signal processing arrangements. It discusses certain physical limitations to the kind of information which can be obtained about a completely unknown far-field distribution, even if noise is assumed to be absent. It is shown, for example, that the "fineness of detail" which can be discerned by means of a given aperture is determined practically by the dimensions of the aperture in terms of the half-wavelength at a frequency corresponding to half of the total frequency bandwidth of the signal waveform. Additive processing of the outputs of the elements of a spatial array can be shown to give the best result when the field is time-stationary but unknown. More complicated processes, such as multiplication and time averaging, have advantages when the field is either non-stationary in the time domain or when prior information about its form is available (e.g. knowledge that it is due to a limited number of "point" sources rather than a continuous source distribution). It is shown that, for a field at a single frequency W c/s, a multiplicative ('correlation') process provides the same fineness of detail of the far-field as that obtainable with an additive array of twice the length, provided the field has a time-stationary amplitude distribution and a relatively rapidly time-varying phase distribution. Furthermore, provided the field is of this type, comparable performance can be obtained by retaining only two receiving elements, situated at the extremities of the aperture, and using a 'multi-frequency' signal waveform whose spectrum occupies a total bandwidth $2W$ c/s.

1. Introduction

A general approach to the angular distribution of the far-field of a directional receiving array can be made without discussing any particular method of signal processing, the object being to establish the limiting performance which could be expected if an ideal system were available. Any proposed system can then be examined, in the light of this ideal, in order to determine how closely its performance approaches to the optimum. Having settled this point, the further inevitable degradation of performance due to the presence of noise in any real system, can be discussed as a separate issue.

2. Statement of Problem

The problem to be studied can be stated briefly as follows. Choose a point in the vicinity of the observer and describe an imaginary sphere with this point as centre and a radius which is very large compared with the overall dimensions of any receiving array likely to be used. Suppose that, over the surface of the sphere, there is a distribution of point sources, each emitting a sinusoidally-oscillating field of frequency ω. Denote the amplitude and relative phase of each resulting field component, as measured at the centre of the sphere, by a complex function $Q(\theta, \phi)$ where θ and ϕ are the polar co-ordinate angles of the corresponding source. The function $Q(\theta, \phi)$ is known as the far-field polar distribution function. In general, $Q(\theta, \phi)$ may also be a function of the range r and of time t. More will be said about this aspect later, but for the present it will be assumed that we are dealing with the particular case of a stationary field in which ω and r are constant and $Q(\theta, \phi)$ is independent of t.

It can be shown[1] that a field of this type can also be completely defined in terms of a complex function $P(y, z)$ which describes the amplitude and relative phase of the field at every point in an infinite reference plane which contains the centre of the imaginary sphere and will be referred to as the "aperture" plane. Since the field is defined equally well by $Q(\theta, \phi)$ or $P(y, z)$, it is clear that there must be a close relationship between these two functions and also that a knowledge of $P(y, z)$ for all values of y

† Electrical Engineering Department, University of Birmingham.

and z should enable $Q(\theta, \phi)$ to be determined completely. This is, of course, the basic idea underlying the whole theory of planar directional receiving arrays.

Although the present study will be restricted to planar apertures, it is perhaps worth while drawing attention to the fact that the aperture function $P(y, z)$ could equally well have been defined over any other closed surface (the infinite plane being regarded as the limiting case of a closed surface whose radius of curvature has been made to tend to infinity). For example, $P(y, z)$ could be defined over a cylindrical surface of infinite axial extent, thus giving rise to a theory based on apertures whose cross-section is the whole or part of an ellipse or a circle. The theory relating to curved apertures of this type will form the subject of a separate paper.

A brief mention must also be made here of an alternative approach to the problem, used for example by Stocklin,[2] in which the field is defined in terms of a three-dimensional function representing its observed values throughout a finite volume rather than over the closed surface which encloses the volume. The two approaches are related by Green's theorem and both must lead to the same kind of conclusions about the amount of directional information which can be received by an array of given physical dimensions. The difference is that, in one case, the "aperture" must be imagined as a finite volume with receivers distributed throughout it and, in the other, either as a finite "window" in an infinite plane.

The far-field can be resolved into two mutually-perpendicular component fields, one of which has constant amplitude and phase in the y-direction and the other in the z-direction. The aperture-plane distribution can be represented by two complex functions $P(y, 0)$ and $R(0, z)$, while the polar distribution is represented by two complex functions $Q(\theta, 0)$ and $S(0, \phi)$.

For the present purpose we shall consider only one of these two-dimensional fields, that is the relationship between a far-field distribution around a circle centred at the origin and the distribution along an aperture line passing through the origin.

The problem is to use an aperture line of finite length to obtain as much information as possible about the far-field; in other words, to use a knowledge of $P(y)$ over a limited range of y in order to obtain some information about the probable form of $Q(\theta)$ in the range $-\pi < \theta < \pi$.

It is convenient to re-define P and Q, using ky where $k = 2\pi/\lambda = \omega/c$ and s, where $s = \sin \theta$, in place of y and θ, respectively. When this is done, it can be shown quite easily that $P(ky)$ and $Q(s)$ are a pair of Fourier Transforms.

3. Computation of Far-field Distribution from Observed Values of Aperture Function

The task of the receiving system is to determine the unknown far-field distribution $Q(s)$ from measurements of $P(ky)$ in the aperture plane. Although $Q(s)$ can only exist over the finite range $-1 < s < 1$, there is no such restriction on $P(ky)$. In general therefore, it would be necessary to know $P(ky)$ for all values of ky from $-\infty$ to ∞ in order to be able to determine $Q(s)$ exactly. Since measurements can only be made over a finite range $-kl/2 < ky < kl/2$ where l is the length of the aperture, it follows that the function $P(ky)$ will generally not be completely known† so that its Fourier Transform $Q(s)$ cannot be computed exactly from the observed data. In other words, the exact form of an unknown far-field, as seen through a "window" of length l, will be uncertain, the degree of uncertainty approaching zero as kl approaches infinity but increasing progressively as kl is reduced.

The observed version of $P(ky)$ can be expressed in the form

$$P_1(ky) = T(ky).P(ky)$$

with a corresponding transform

$$Q_1(s) = D(s).Q(s) \qquad \ldots\ldots(1)$$

where $D(s)$ is the usual directional-pattern function given by the Fourier Transform of the aperture sensitivity function $T(ky)$. If it is assumed that $T(ky) = 0$ for $|ky| > kl/2$ (i.e. that no information is available about the behaviour of $P(ky)$ outside the limit of an effective aperture of length l) then it can be shown (see Appendix) that the reconstructed version $Q_1(s)$ of the far-field function $Q(s)$ can be represented exactly by discrete samples in the s-domain at intervals of λ/l. The smoothed form of $Q_1(s)$ can be synthesized by multiplying each of these complex samples of $Q_1(s)$ by a directional "weighting" function $D(s)$. Note that the samples should extend along the whole s-axis and that some further uncertainty must occur as a result of the omission of all samples outside the range $-1 < s < 1$. This effect ('truncation error') will be most noticeable near the edges of the s-spectrum. There are two distinct kinds of uncertainty about the exact form of the far-field; one due to the restricted range of ky over which measurements are possible and one due to the fact that the spectrum itself is "band-limited" by the physical restriction of the far-field directional spectrum to the range $-1 < s < 1$. Both of these uncertainties tend to zero as the aperture length is increased indefinitely but they may become of great practical significance if the aperture length is reduced until it corresponds to only a few half-wavelengths.

† See Appendix.

The "taper", "sensitivity" or "weighting" function (all these terms will be found, in various contexts, in the relevant literature) need not be restricted to a rectangular shape (referred to by some writers as a "boxcar" function). Any shape is suitable provided the function is zero for $|ky| > kl/2$.

A great deal of attention has been paid in the literature on array theory, to the question of the optimum choice of $T(ky)$, and therefore of the directional "pattern" denoted by its transform $D(s)$, in order to obtain some real or fancied improvement in the agreement between the reconstructed far-field denoted by $Q_1(s)$ and its real form denoted by $Q(s)$. It does not always seem to have been properly understood that the answer to this question will be greatly influenced if, as is almost invariably the case, some prior knowledge is available about the probable form of $Q(s)$; if, for example, it is known that a particular $Q(s)$ happens to be produced by a small number of separate sources rather than by a continuous source distribution.

The significance of prior information about $Q(s)$ can be explained in the following way. The reconstructed version of $Q(s)$ is defined by a set of ordinates spaced λ/l apart over the range $-1 < s < 1$. Since, assuming that l is an integral multiple of $\lambda/2$, there will be n such ordinates in the range, where $n = 2l/\lambda$, the reconstructed version of the far-field has n (and only n) complex degrees of freedom. Suppose now that it is known in advance that $Q(s)$ is in fact due to a number m of point sources, so that $Q(s)$ consists of a corresponding number of spectral "lines" $Q(s_1)$, $Q(s_2)$, $Q(s_3)$ etc. The problem is to find both the positions and the values of these spectral components. Since $Q(s)$ is complex however, each component has three degrees of freedom, (i.e. amplitude, phase and bearing) so that it would appear that the maximum number of sources which can be identified will be obtained by choosing the largest value of m for which $3m \leqslant 2n$. Usually however we are interested only in the modulus of $Q(s)$. In this case, the amplitude and bearing of each source represents only two degrees of freedom and $m \leqslant n$; that is, in the assumed absence of noise, up to n separate point sources can be identified exactly in magnitude and bearing, where n is the number of half-wavelengths contained in the length of the aperture.

The interesting point here is that this is true, irrespective of the positions of the sources; in the absence of noise it is theoretically possible to resolve sources of any magnitude, no matter how close together any adjacant pair may be, provided the number of sources does not exceed the number of half-wavelengths in the aperture. This means that care must be taken in defining what is meant by the "angular resolution" of an array. It has sometimes been taken to refer to the closest approach at which two point sources can be identified when it is known in advance that only these two sources are present. The theoretical maximum "resolution" under these conditions is infinite, in the absence of noise, and some other criterion is necessary in order to define the performance of a receiving system. If the far-field is unknown, the term "fineness of detail" is a better one than "resolution".

4. Sampling of the Aperture Function

Suppose next that the "taper" function $T(ky)$, with which the actual distribution $P(ky)$ must be multiplied to obtain its observed form, is not only restricted to the finite range $-kl/2 < ky < kl/2$ but consists only of discrete ordinates spaced equally within the aperture range. Then, if this spacing is denoted by ky_1, the directional function $D(s)$ will be periodic with a period of $2\pi/ky_1 = \lambda/y_1$ in the s-domain, so that the observed far-field $Q_1(s)$ will consist of $Q(s)$ convolved with a periodic function. This will yield results bearing a reasonable resemblance to the actual field $Q(s)$ provided the adjacent peaks of the periodic directional pattern always fall outside the range $-1 < s < 1$, even when $D(s)$ has a peak centred at either of the limits $s = -1$ and $s = 1$. $Q_1(s)$ is likely to differ widely from $Q(s)$ if this condition is not met, that is, unless the period of $D(s)$ in the s-domain is at least 2; i.e.

$$\frac{\lambda}{y_1} \leqslant 2 \quad \text{or} \quad y_1 \leqslant \frac{\lambda}{2}$$

This is analogous to the "sampling" theorem in communication theory, which states that the frequency spectrum of a band-limited time-waveform (and hence the waveform itself) is completely defined by samples taken at intervals, provided their time spacing does not exceed the reciprocal of twice the positive-frequency bandwidth (i.e. $\frac{1}{2}\pi$ times the reciprocal of the complete bandwidth in the W domain). The theorem is often misunderstood and taken to mean that the waveform over the period of observation is completely defined by samples taken *only over that interval*. This is not true; the waveform over the observation period is completely defined by samples at intervals of $\frac{1}{2}W$ (where the frequency band extends from $\omega = 0$ to $\omega = 2\pi W$), *taken over all time*. The sampling period should extend appreciably beyond the period over which the waveform is to be reconstructed.[13] Provided $WT \gg 1$ as it usually is in the communication context, this extension of the sampling period beyond the observation period forms only a small percentage of the period and thus the uncertainty in the reconstructed waveform, caused by ignoring all samples outside the reconstruction period, is small. Care must be taken not to extend arguments, which are very nearly true when $WT \gg 1$, down to cases where WT is allowed to approach unity.

In the array case, the effect of aperture sampling instead of continuous measurement of the aperture function is simply to change the form of the directional "weighting" function with which $Q(s)$ must be convolved; it becomes periodic. Provided the product $2kl \gg 1$ (i.e. the aperture contains many wavelengths) this change will not be significant.

This is perhaps an appropriate place to point out that, in many practical applications, a particular type of prior information is available, namely the knowledge that the far-field function $Q(s)$ is effectively zero (possibly due to inherent directivity of the aperture sampling devices) except over a range $-s_{max} < s < s_{max}$ where $s_{max} < 1$; in other words, the field distribution is confined to an angular sector $\pm \sin^{-1}(s_{max})$. The spacing of the samples in the s-domain remains λ/l so that the number of samples is reduced by a factor s_{max} and the minimum spacing of the aperture samples is correspondingly increased from $\lambda/2$ to $\lambda/2 \,.\, 1/s_{max}$.

To sum up, assuming that $P(ky)$ is completely unknown for $|ky| < kl/2$, the apparent far-field function $Q_1(s)$ is completely defined by complex ordinates spaced λ/l apart on the s-axis and, provided n is large, it is very nearly defined by the n such samples which fall within the range $-1 < s < 1$. It seems reasonable to find that all the available information contained in the part of $P(ky)$ falling within the aperture can be represented by an equal number of complex samples of $P(ky)$ spaced $\lambda/2$ apart within the available range of ky.

This leads to an interesting concept of the aperture, together with its Fourier Transform computer, as a potential image-forming device which is analogous to an optical lens. The dependence of the "fineness of detail" or "definition" of the far-field as "seen" through the aperture is closely analogous to the dependence of the definition of an optical system (e.g. a microscope) on the number of wavelengths contained in its apparent aperture.[6]

The function $Q_1(s)$ can be said to have a "spectrum" in the ky domain which exists only over the range $-kl/2 < ky < kl/2$. The squared-modulus $|Q_1(s)|^2$ will have a ky "spectrum" of exactly twice this width so that, by the sampling theorem, $|Q_1(s)|^2$ will be represented by *real* ordinates of $|Q_1(s)|^2$ spaced $\lambda/2l$ apart and will be very nearly represented by the $2n$ such samples which fall in the range $-1 < s < 1$. This shows that the available $2n$ degree of freedom can be used to define $|Q_1(s)|^2$ directly if we wish, without any need to know the phase of $Q_1(s)$. The far-field, as "seen" through the aperture, will have a fineness of detail which can be thought of either as due to n complex samples spaced $\lambda/2$ apart or to $2n$ "real" samples spaced $\lambda/2l$ apart.

5. Formation of Fourier Transform by "Scanning" Process

The process of forming the Fourier Transform of the observed aperture function $P(ky) \,.\, T(ky)$ can be interpreted as that of forming the additive directional pattern $D(s)$ corresponding to the taper function $T(ky)$ and then using this to "scan" the far-field.[7] The resultant apparent far-field is interpreted as being the result of scanning the true far-field with an imperfect directional pattern having a finite "beamwidth". This way of visualizing the process of convolution and transformation is, of course, quite sound provided the hypothetical "beams" are not taken too literally.

It is true that the fineness of detail of the apparent far-field is represented by a limited number of ordinates on the s-scale and it is also true that the main lobe of the directional pattern has a definite shape in the s-domain so that its beamwidth can be defined in terms of the spacing of these ordinates. It is however dangerous to attempt to draw general conclusions about the relationship between "angular resolution" and "beamwidth".

6. Multiplicative (Time-Average-Product) Arrays

Since it has been shown that an additive process enables all the available information about an unknown, time-stationary, single-frequency field to be extracted from a given aperture, it follows that nothing is to be gained by any other process for a field of this type. It can be shown however that another process, involving multiplication and time-averaging of aperture samples, does offer certain advantages in some cases where some or all of these conditions are not met. The theory of multiplicative aperture processing is dealt with elsewhere[4, 5] but the main points will be summarized briefly here.

The aperture is split into two parts with their centres separated by a distance y_0 and the time-average of the product of the total outputs from the two parts of the array is formed. Using suffixes 1 and 2 to denote the two parts of the aperture, the apparent field produced by a single point source on a bearing s_a will be of the form

$$\begin{aligned} f(s) &= |Q(s_a)|^2 \,.\, D_1(s-s_a)D_2(s-s_a)\cos ky_0(s-s_a) \\ &= |Q(s_a)|^2 \,.\, D(s-s_a) \end{aligned}$$

This can be shown to have a main lobe which has only half the width, between zeros, of the directional pattern for simple additive processing of the aperture. Now suppose that the field is known to be due to two point sources on bearings s_a and s_b, respectively.

The response will be of the form

$$|Q(s_a)|^2 . D(s-s_a) + |Q(s_b)|^2 . D(s-s_b) + \\ + |Q(s_a) . Q(s_b)| . D_2(s-s_a) D_1(s-s_b) \cos\alpha \\ + |Q(s_a) . Q(s_b)| . D_1(s-s_a) D_2(s-s_b) \cos\beta$$

where $D(s) = D_1(s) D_2(s) \cos ky_0 s$ and α and β depend on the phase angles of the signals from the two sources. The first two terms are the "wanted" ones which correspond to the convolution of $|Q(s)|^2$ and $D(s)$ but the second two are cross-product terms which convey no useful information and, in general, will tend to obscure the wanted information conveyed by the first two terms.

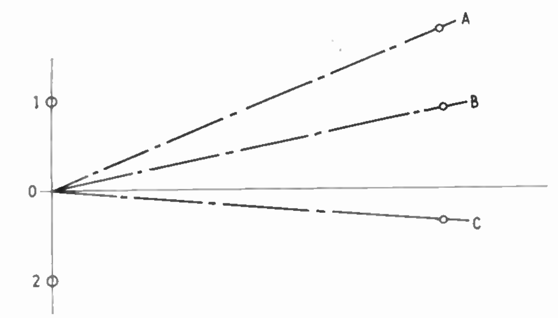
Suppose however that the field is not time-stationary. Then, provided the intensity and bearings of the two sources remain sensibly constant while a relatively rapid variation of the phase angles of the two signals takes place, the "wanted" terms will remain steady while the "unwanted" ones will vary. If the output is averaged over a sufficiently long time, the "unwanted" effects will be reduced to a negligibly small value.

For m separate point sources, the number of cross-products will be $m(m-1)$ so that, assuming roughly the same power constant for each, the r.m.s. ratio of "unwanted" to "wanted" products will increase roughly proportionally with m. This means that, as the number of point sources is increased, the averaging time, necessary to reduce the "noise" which they represent to an acceptable value, will increase roughly in proportion. Also, of course, it will be necessary for the relative rate of change of phase, compared with the rate of change of amplitude or bearing angle, to increase correspondingly if the unwanted products are to be reduced to a negligible value without appreciably changing the wanted signals.

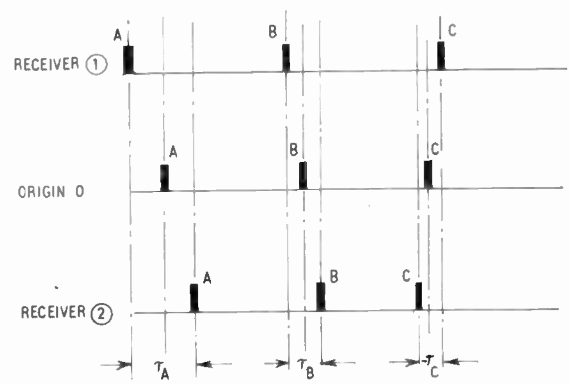
The conclusion reached is that, under suitable conditions, multiplication is capable of providing $2n$ real samples of the modulus of $Q(s)$ instead of the n complex samples of $Q(s)$ given by additive methods. Since the recovery of the complex samples would require a process of synchronous demodulation which is usually not possible in practice, only the modulus of $Q(s)$ is likely to be available in an additive system anyway. Thus multiplication can be said to double the number of samples of $|Q(s)|$ and to give a fineness of detail corresponding to that obtainable with an additive array of twice the physical length.

7. Non-Sinusoidal Fields

Suppose now that $Q(s)$ represents not the amplitude and phase of a single sinusoidal time-function but the amplitude and time-position of some other more complicated repetitive function of time. The far-field can be imagined as being due to a distribution of



(a) Two receivers 1 and 2; three sources A, B and C on different bearings.



(b) Diagrams showing time-waveforms of signals at the two receivers. The time difference τ_A, τ_B, τ_C, provided are measurements of the bearings of sources A, B and C provided no "overlapping" occurs. (Note that the reversed sign of τ_C indicates that source C has a negative bearing angle.)

Fig. 1. Relationship between the resulting waveform at the origin and at each of two receiving elements in the aperture plane.

point sources, each emitting the same waveform except for differences of amplitude and time-phase. Thus, if the signal waveform is denoted by $f(t)$, then the signal measured at the origin, from a far-field source denoted by $Q(s)$, will be $|Q(s)| \times f[t - t_1(x)]$

where $$Q(s) = |Q(s)| \exp j \left[\frac{2\pi}{T} t_1(x)\right]$$

and T is the repetition period.

Some general properties of a field of this nature can be demonstrated by assigning a particular form of $f(t)$, namely that of a series of short, equi-spaced pulses of constant amplitude. There is really no lack of generality involved here because, in principle, any other *known* waveform could be converted into a train of pulses with the same repetition period merely by the use of suitable passive networks at the outputs of the array sampling points. Suppose, for example, that there happen to be three point sources A, B and C.

all of the same intensity but on different directional bearings, emitting signals of this type. Figure 1 illustrates the relationship between the resulting waveform at the origin and at each of two receiving elements in the aperture plane at equal distances on either side of the origin. It can be seen that, for waveforms as drawn in Fig. 1, it would be possible to determine the bearings of the individual sources unambiguously by measuring the time differences between the various pairs of pulses received at the two elements. The sign of the bearing angle is implied in each case by the algebraic sign of the difference and the pulses which are closest together are to be classed as pairs. The condition for an unambiguous result is that the gaps between the pairs of pulses must always be greater than the maximum separation of the pulses forming each pair. Although the diagrams have been drawn for only three sources, the argument can evidently be extended to any number of sources and there is no need for the amplitudes to be assumed equal.

It can be concluded that resolution of the separate sources will be possible if, and only if, the time-separation of the source waveforms is known to exceed a certain minimum value. This is a situation which would be most unlikely to occur in practice, except perhaps in the case of some kinds of navigational aids in which the relative time-phases of the signals from several fixed transmitters are controlled from a common source.

Suppose however that the directional amplitude distribution of the sources remains constant but that their time-phases are allowed to vary with time. It can be seen fairly easily that such a variation will, in principle, enable the ambiguities to be removed, even if the condition relating to the time-separation of the source waveforms is relaxed. Correct pairing of the pulses received at the two receivers will produce constant time-differences corresponding to correct source bearings, while spurious "sources", due to incorrect pairing of pulses, can be ignored because the bearing indications will not be constant. Under these conditions, the ultimate limit of directional resolution will be set by the length of the pulses forming the signal waveform. Since the minimum length of these pulses is fixed by the available signal frequency bandwidth, it follows that, as might be expected, the angular resolution, or "fineness of detail" which can be obtained by means of two receivers at a given distance apart, will depend on the signal frequency bandwidth.

Assuming that the signal spectrum is uniform and extends from 0 to W c/s, the pulse time-waveform will be

$$f(t) = \frac{\sin 2\pi Wt}{2\pi Wt}$$

giving a pulse length, between zeros, of $1/W$ seconds. This will represent the minimum separation at which a pair of pulses can be resolved as separate entities, irrespective of their relative amplitudes.

Since there are only WT such intervals within the repetition period T, it follows that the maximum number of separate sources which can be resolved will also be WT. If these sources are imagined to be distributed evenly over the range $-1 < s < 1$ in the s-domain their spacing will be $2/WT$. For two receivers spaced l apart, the time differences for sources on bearings s_1, s_2 respectively will differ by $l(s_2 - s_1)/c$. This difference must be less than the pulse length $1/W$.

$$l\frac{(s_1 - s_2)}{c} \nless \frac{1}{W}$$

i.e.

$$(s_2 - s_1) \nless \frac{c}{Wl}$$

Thus giving a maximum of $2Wl/c$ sources in the range $-1 < s < 1$.

This result is of basic importance. It shows that the maximum number of separate sources (each emitting the same waveform), which can be resolved by means of an aperture of length l and a frequency bandwidth $0 < f < W$, is proportional to the product of the aperture and the bandwidth. Furthermore, all available directional information, (always in the assumed absence of noise) can be obtained by means of only two receivers, placed at the extremities of the aperture. The result has been obtained without any discussion of the actual processing which is necessary to enable the various time-differences to be measured. It would not be surprising to find that there are, in fact, a number of methods which could be used and which could be examined to determine to what extent they realize the performance to be expected from an ideal optimum system.

Since the directional resolution is proportional to Wl, the aperture can be thought of as having a two-dimensional existence in the space-frequency domain where it possesses an "area" of Wl.

As mentioned earlier, the fact that the result has been based on the assumption of a pulse-type signal waveform does not impair its generality. It will, in fact, be shown (see Sect. 7.1) that it is not absolutely necessary to know the shape of the time-waveform at all provided the modulus of its complex frequency spectrum is known. This implies, of course, that the time-waveforms of the point sources can all have the same "power" frequency spectrum and that this spectrum is known in advance.

The frequency spectrum of any periodic time-waveform with a period T can be represented approximately by a series of WT complex spectral "lines" spaced $2\pi/T$ apart. Thus the number of sources which can be distinguished, by means of two receivers at the extremities of the aperture, is just equal to the minimum number of frequency components which are needed to synthesize the signal waveform.

It is interesting to compare this result with the "resolution" obtainable by retaining only a single frequency W and an aperture l but using an array of receiving elements at half-wavelength spacing in the aperture. The maximum number of sources which can be distinguished is equal to the number of half-wavelengths in l, which is again equal to $2Wl/c$.

It can be concluded that a two-element array, using a multi-frequency (i.e. repetitive) signal waveform, is potentially capable of providing the same amount of directional detail of the far-field as a single-frequency array of the same length, working at the highest frequency within the same frequency band. This can only be done however if the condition of time-stationary amplitude and time-varying phase distribution is satisfactorily met.

So far it has been assumed that the frequency band extends from 0 to W c/s. The band W c/s wide could be translated, by the usual method of amplitude modulation and filtration, to any other position in the spectrum. In order to regain the original signal waveform however, synchronous demodulation with the original carrier would be necessary. This is often impracticable and can be avoided by transmitting the complete modulated carrier waveform. This requires a total bandwidth of $\Delta f = 2W$ c/s. It can be stated generally therefore that a total bandwidth Δf and a total aperture l will enable $l\Delta f/c$ separate sources to be resolved. This number would be doubled if the synchronous demodulation could be arranged.

7.1. *Methods of Measuring Time-differences between Waveforms*

The processing of the outputs of two receivers to obtain directional information is essentially concerned with the comparison of time-waveforms in order to determine time-displacements of their various components. Take, for example, the simplest case of two identical waveforms, each consisting of a train of short equally spaced pulses. An obvious method of measuring the time-displacement would be to photograph both waveforms and then to slide one over the other until coincidence is observed. The same result could be achieved by applying a known, variable time-delay to the appropriate one of the pair of waveforms and then using a multiplier circuit to indicate when coincidence occurs; the multiplier will give an output if, and only if, the pulses overlap. It is difficult to avoid the view that all methods of measuring time-differences involve, either directly or indirectly, the application of a known, variable time delay to one of the waveforms, followed by a process of multiplication. The result then appears when the behaviour of the product, as a function of the applied time-delay. is examined in a suitable way. The word "correlation" has been carefully avoided so far, but it is easy to see the connection between the process of measuring the unknown time-difference between two waveforms on one hand, and the mathematical concept of a correlation function on the other. Confusion may arise here however because the original mathematical concept was concerned with the average of the product, taken over all time, rather than directly with the product itself.

Nevertheless, when the time-function $f(t)$ is periodic, all possible information about it is contained in each period T, and the behaviour of the time-average of the product $f(t) \, . \, f(t-\tau)$, is exactly the same, when the averaging time is T, as it would be over infinite time. Thus, for a periodic waveform denoted by $f(t)$, the correlation function $\rho(\tau)$ can be completely defined by observation over the period T. Use can then be made of a well-known theorem which states that $\rho(\tau)$ is the Fourier Transform of the squared-modulus of the complex frequency spectrum of $f(t)$. In particular, if the spectrum is uniform over the band 0 to W c/s, then

$$\rho(\tau) = \frac{\sin \pi W \tau}{\pi W \tau}$$

It will be noted that only the modulus of the frequency spectrum of $f(t)$ is concerned and not the actual time-waveform.

The process of locating the sources by observing the form of $\rho(\tau)$ can be visualized as one of "scanning" by means of a "beam pattern" $\sin 2TWt/\pi Wt$. This has a "beamwidth" (between zeros) of $1/2W$ in the τ-domain, thus leading once again to the conclusion that $2Wl/c$ separate sources can be resolved.

We have thus proved that "correlation" by the use of a time-varying delay, followed by multiplication and averaging, is not only a possible method of determining the directional distribution of the far-field but it is also an optimum one, in the sense that it enables all available directional information to be extracted.[9, 10]

An alternative method of processing is that described by Kay,[8] in which the directional information is obtained by analysing the frequency spectrum of the product $f(t) \, . \, f(t-\tau)$. For this purpose it is necessary to use a particular kind of time-waveform, namely a sinusoidal carrier which is frequency-modulated by a linear "sawtooth" function so that the maximum frequency deviation has a value which will be denoted by $\pm\delta\omega$. It can be shown[11] that the envelope of the frequency spectrum of $f(t) \,.\, f(t-\tau)$ in this particular case, contains two components, each with a $\sin x/x$ type of profile, centred at $\omega = 2\delta\omega\tau/T$ and $\omega = 2\delta\omega(1-\tau/T)$, respectively, and having relative maximum amplitudes of $(1-\tau/T)$ and τ/T, respectively.

Although the two parts of the spectrum overlap, it is found that, provided $\tau/T \ll 1$, the first part will predominate and the envelope will then be represented by a function of the form $\sin \frac{1}{2}\omega\tau/\frac{1}{2}\omega\tau$, centred at the "beat frequency" $2\delta\omega\tau/T$. Sources on different bearings will produce similar spectral envelopes centred at different values of $2\delta\omega\tau/T$. The limit of resolution for sources on bearings s_1 and s_2, producing time-differences τ_1 and τ_2, respectively, will occur when the difference between the beat frequencies ω_1 and ω_2 is such that

$$\frac{\omega_2-\omega_1}{2} = 2\pi$$

i.e.

$$\tau_2-\tau_1 = \frac{2\pi}{\delta\omega}$$

or

$$s_2-s_1 = \frac{2\pi c}{l\delta\omega}$$

The maximum number of sources, at this spacing, in the range $-1 < s < 1$, is $l\delta\omega/\pi c$.

Finally, provided T is large compared with $2\pi/\delta\omega$, (i.e. the modulation index of the frequency modulation is large), the total bandwidth Δf will be very nearly equal to $2\delta\omega/2\pi$, so that the maximum number of sources can be expressed as $l\Delta f/c$.

This result shows that this method of processing can also give a directional resolution which approaches the optimum. It suffers from two disadvantages however. Firstly, it is necessary to use a particular kind of transmitted waveform (and to ensure, in an echo-ranging system for example, that this waveform is not destroyed on reflection from the target), whereas the same information could theoretically have been obtained by means of any waveform having the same power spectrum. Secondly, the periodic time must be large compared with $1/\Delta f$. This implies that the observation has to be extended over a period which is considerably longer than the minimum period required for optimum processing, or the "lines" of the frequency spectrum have to be placed much more closely together than is really necessary. On the other hand, the "f.m. scan" method has the advantage that it avoids the necessity for providing some means of applying the time-varying delay which is an essential part of the "correlation" method.

8. Conclusions

It has been shown that the ability of a given aperture to provide information about the angular distribution of a single-frequency far-field is limited, even in the complete absence of noise. The reason for this is fundamental to the wave nature of the field and is closely analogous to the problem of forming an optical image of a distant scene. It is well known that the fineness of detail which can be observed is limited, even for a perfect medium and in the absence of noise, and it can only be increased by reducing the wavelength of the radiation used.

A distinction must be drawn between "angular resolution" of point sources and the "fineness of detail" of an unknown distribution. There is no limit to the possible accuracy of angular resolution of point sources provided it is known in advance that the field is due to discrete point sources and provided their number does not exceed a limiting value which depends on the length of the aperture. In the absence of such prior information it is physically impossible, using an aperture of finite length, to determine whether the observed field is due to discrete point sources or to some equivalent continuous distribution. The apparent angular resolution of a given practical receiving system, when measured for the field due to two known sources, must not be taken as necessarily giving a reliable measure of the "fineness of detail" in the image which it produces of an unknown far-field.

The maximum number of point sources which can be completely identified, in amplitude and phase, in a single-frequency far-field, is equal to the number of half-wavelengths contained in the length of the aperture. This number can be doubled if only the amplitudes of the sources are of interest, provided their relative phases are known to vary rapidly with time while their angular distribution remains relatively constant.

If, and only if, the latter condition is met, only two receiving elements at the extremities of the aperture are needed, when used in conjunction with a non-sinusoidal signal waveform having a frequency bandwidth represented by $f_m \pm \delta f$ where f_m is the mid-band frequency. The maximum number of separate sources which can be resolved is then $2l\delta f/c$ where l is the length of the aperture and c is the velocity of propagation in the medium. This number can also be expressed as the number of half-wavelengths in the aperture at the frequency δf.

In general, if a "correlation" method of processing is used only the "power" frequency spectrum of the signal need be known and not its actual time-waveform. An alternative method ("f.m. scan") avoids the necessity for a time-varying delay device in the processing circuits but is restricted to a particular kind of time-waveform.

9. References

1. P. M. Woodward and J. D. Lawson, "The theoretical precision with which a radiation pattern may be obtained from a source of finite size", *J. Instn Elect. Engrs*, **95**, p. 363, 1948.
2. P. L. Stocklin, "Space-time decision theory of acoustic field processing". Paper read at Underwater Acoustics Symposium at Imperial College, London, August, 1961. Published in "Underwater Acoustics", ed. V. M. Albers, p. 339 (Plenum Press, New York, 1963).
3. R. B. Blackman and J. W. Tukey, "The Measurement of Power Spectra" (Dover, New York, 1958).
4. V. G. Welsby and D. G. Tucker, "Multiplicative receiving arrays", *J. Brit.I.R.E.*, **19**, p. 369, June 1959.
5. V. G. Welsby, "Multiplicative receiving arrays: the angular resolution of targets in a sonar system with electronic scanning", *J. Brit.I.R.E.*, **22**, p. 5, July 1961.
6. G. Toraldo Di Francia, "Resolving power and information", *J. Opt. Soc. Amer.*, **45**, p. 497, July 1955.
7. D. G. Tucker, "An analogue computer for Fourier transforms", *J. Brit.I.R.E.*, **18**, p. 233, April 1958.
8. L. Kay, "A plausible explanation of the bat's echo-location acuity", *Animal Behaviour*, **10**, Nos. 1 & 2, p. 34, January–April 1962.
9. V. G. Welsby, "A two-element aerial array", *Electronic Technology*, **38**, p. 160, May 1961.
10. V. G. Welsby, "Electronic sector-scanning array", *Electronic Technology*, **39**, p. 13, January 1962.
11. A. J. Hymans and J. Lait, "Analysis of an f.m. continuous-wave ranging system", *Proc. Instn Elect. Engrs*, **107B**, p. 365, July 1960.
12. G. Toraldo Di Francia, "Directivity, supergain and information", *Trans. Inst. Radio Engrs (Antennas and Propagation)*, **AP-4**, p. 473, July 1956.
13. H. D. Helms and J. B. Thomas, "Truncation error of sampling theorem expansions", *Proc. Instn Radio Engrs*, **50**, p. 179, February 1962.
14. A. Ksienski, "Spatial frequency characteristics of ferrite aperture antennas", Paper read at Symposium on "Electromagnetic Theory and Antennae", Copenhagen, June 1962.
15. R. M. Chisholm, "Transfer functions and resolving power of radio telescopes". Paper read at Symposium on "Electromagnetic Theory and Antennae", Copenhagen, June 1962.

10. Appendix: Effective Length of Aperture

Various authors[14, 15] have pointed out that the effective range $-kl/2 < ky < kl/2$ over which $P(ky)$ is known, may appreciably exceed that corresponding to the physical length of the aperture; the reason for this may be summarized as follows.

Since $Q(s)$ cannot exist outside the range $-1 < s < 1$ it can be proved rigorously that $P(ky)$ can be represented *exactly* over the range $-kl/2 < ky < kl/2$ by samples of $P(ky)$ taken at intervals of π (i.e. at intervals of $\lambda/2$ along the reference plane) *throughout the whole range* $-\infty < y < \infty$. The so-called "truncation" error occurs because of the omission of all samples outside the range $-kl/2 < ky < kl/2$. The error is naturally greatest near the edges of the physical aperture and becomes progressively more serious as the aperture length is reduced in half-wavelengths. It can be reduced by increasing the number of samples, particularly near the edges of the aperture—hence the idea of non-uniform sampling. If a certain degree of accuracy is required, simple half-wavelength sampling may be adequate if the aperture is "long" but rather closer and non-uniform spacing may be necessary if it is "short". The relevant point here is that, for finite accuracy of reconstruction of $P(ky)$ within the aperture, only a finite number of samples is required. This leaves open the possibility of introducing still more sampling points and using them to measure the *derivatives* of $P(ky)$ and hence to reconstruct $P(ky)$ *outside* the physical aperture. (This is the real significance of superdirectivity.)

By applying Taylor's Theorem, the derivatives can be used to reconstruct the function $P(ky)$ at points outside the aperture. For example, if the actual aperture contains n sampling points at half-wavelength intervals, a knowledge of the derivatives of $P(ky)$ at the nth sampling point would enable $P(ky)$ to be reconstructed at a virtual sampling point occurring just outside the physical aperture, and so on. Evidently this process would increase the effective length of the aperture, thus increasing the fineness of detail of the far-field which could be observed.

If one sampling point is placed at $y = l/2$, i.e. at one limit of the physical aperture, the value of $P(ky)$ at any other point $y = l/2 + a$ is such that

$$aP'\left(\frac{kl}{2}\right) + \frac{a^2}{2!}P''\left(\frac{kl}{2}\right) + \frac{a^3}{3!}P'''\left(\frac{kl}{2}\right) = P\left(\frac{kl}{2} + a\right) - P\left(\frac{kl}{2}\right)$$

Measurement of $P(kl/2 + a)$ at m different points within the aperture (i.e. for m different negative values of a), will provide m simultaneous equations which can be solved to give the values of the first m derivatives of $P(ky)$ at the point $kl/2$. This information can then be used to compute $P(ky)$ for some chosen point outside the aperture (i.e. for a chosen positive value of a), provided it is possible to neglect all terms involving higher orders of derivatives than those obtainable from the chosen number of measurement points within the aperture. It is clear that the required number of terms in the Taylor series and thus the number and accuracy of the measurements required, must increase rapidly as the point in question is moved further and further outside the limit of the physical aperture.

So far, on purely theoretical grounds, there appears to be no reason why this process could not be repeated indefinitely, obtaining more and more additional samples of $P(ky)$ at intervals of $\lambda/2$ outside the aperture. This result is based however on the assumption that it is possible to measure $P(ky)$ exactly (or at least to a sufficient degree of accuracy for the purpose). But, quite apart from the question of experimental error, there is also the fundamental fact that, in order to measure a sample of $P(ky)$, it is necessary to make use of some kind of physical transducer. The transducer must have finite dimensions and so cannot really provide a "point" sample. Furthermore, since the transducer must necessarily be coupled in some way to the medium, it is bound to distort the field to a greater or lesser extent so that no exact measurement is possible unless the influence of the transducer on the field can be computed and taken into account.

Manuscript first received by the Institution on 14th June 1962 and in final form on 4th February 1963. (Paper No. 835/SS23)

POINTS FROM THE DISCUSSION

Dr. E. J. Risness†: The author suggests (Section 4, para. 4) that it is impossible within a finite aperture to obtain more information about a far-field than can be gained from sample points $\lambda/2$ apart filling the aperture. This appears to neglect the theoretical possibility of non-uniform space sampling, whereby samples taken closer than $\lambda/2$ within an aperture can be used to reconstruct the field outside the aperture and so effectively widen it theoretically, without limit, provided the correct processing is applied. In practice, as Dr. Vanderkulk's paper‡ suggests, practical systems are unlikely to achieve great gains.

Both Professor Tucker and Dr. Welsby point out the difference between "additive" and "multiplicative" array processing. To the casual listener these differences are emphasized by the terminology used—he knows that addition and multiplication are very different operations. I feel this is unfortunate since in many ways the two methods of array processing are very similar. In "additive" processing the array outputs are added together and then passed through some sort of rectifier or detector. If this detector is a square-law device (as it should be ideally, and in any case all detectors are effectively square-law at low signal/noise ratios), then the system effectively adds all array outputs and squares the total, i.e. it forms all instantaneous products between array outputs, including self products. In "multiplicative" processing, the array is split into two halves, each added internally, then multiplied together. This produces all cross-products between pairs of array outputs, one taken from each half array, but ignores cross-products within each half, or self-product terms. This "multiplicative" processing is like "additive" processing but with certain contributory terms left out. In Dr. Welsby's paper, the description of "additive" array processing as forming a "space-average-product", while "multiplicative" processing leads to a "time-average-product", would seem to emphasize still further their differences at the expense of their similarities. Incidentally, the term "time-average-product" has a different meaning here from that used by others.§ Perhaps the author could comment on these points?

† Admiralty Underwater Weapons Establishment, Portland.

‡ W. Vanderkulk, "Optimum processing for acoustic arrays". Paper read at Symposium on "Sonar Systems", Birmingham, July 1962 (To be published).

The Author (*in reply*): The "aperture" is the range of y over which $P(ky)$ is known. It is the number of half-wavelengths in this "aperture" which determines the fineness of detail of the far-field which can be observed. As Dr. Risness points out, the theoretical possibility certainly exists that the effective "aperture", over which $P(ky)$ is known, may exceed the physical aperture. This effect is likely to be of little practical significance in systems which are already highly directional, owing to the large range over which extrapolation is necessary and the high sampling accuracy which this implies.

The term "additive" is used, for want of a better term, to describe a system which is not "multiplicative", i.e. which does not contain a direct multiplier, as opposed to a detector.

I did not mean to imply that additive processing *forms* a space-average-product, but rather that it is *equivalent* to a space-averaging process, in the particular case where only the modulus of $Q(s)$ is of interest.

The term "time-averaged-product" has been interpreted literally as applying to any process involving one or more multipliers with associated smoothing filters.

§ e.g. D. C. Fakley, *J. Acoust. Soc. Amer.*, **31**, p. 1307, 1959.

U.D.C. 534.88

Sea-bed Echo Amplitude Fluctuations Arising from Ship Motion

By

B. K. GAZEY, Ph.D.†

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

Summary: Sea-bed echo amplitude fluctuations result from a displacement of the transducers of a ship-borne echo-sounder relative to the sea bottom. The origins of these fluctuations are discussed and analysis is given which enables an estimate to be made of the magnitude of the various components of a vessel's motion necessary to ensure that successive echoes are uncorrelated.

It is shown that: (1) horizontal motion is of more importance than vertical motion; (2) randomizing displacements are small compared with the dimensions of the insonified area of the sea-bed; (3) some large discrete scatters must be present if small motion fluctuations are to be experienced. Experimental evidence, obtained using a scale model of an echo-sounding environment, is presented in support of the theoretical conclusions.

1. Introduction

The carrier envelope of an echo-pulse reflected by the sea-bed and displayed on an echo-sounder can be observed to fluctuate violently in amplitude and shape when the vessel carrying the sounding equipment is both under way and, surprisingly, at anchor. The order of magnitude of these fluctuations is shown by the consecutive echo-sounder traces reproduced in Fig. 1(*a*) in which the interval between transmissions was 4 seconds and the depth of the sea-bed was 14 fathoms. These are typical samples taken from a large series of traces recorded whilst steaming over a dredged channel, and later, whilst at anchor off Falmouth during underwater acoustic propagation studies aboard the R.R.S. *Discovery II* in October 1960.

The fluctuations observed when the vessel was at anchor were too large to be attributed solely to forward scattering of the acoustic beam by the thermal inhomogeneities within the medium.[1–4] This is demonstrated by the almost complete absence of fluctuations of the echoes reflected from an 18 in diameter, hollow, spherical target, positioned one fathom above the bottom; these are recorded on the same trace as the fluctuating bottom echo (Fig. 1(*b*)).

The fluctuations must, therefore, be due to the three residual motions normally possessed by a vessel even when at free anchorage, i.e. pitching, rolling and yawing.

† Electrical Engineering Department, University of Birmingham.

2. The Causes of Amplitude Fluctuations

Apart from such factors as the changing thermal microstructure of the ocean and variations in the coupling between the transducers and the medium, the amplitude of the echo received from the sea-bed is dependent upon the following:

(a) the positions of the transmitting and receiving transducers relative to the bottom, (i.e. the geometrical configuration),

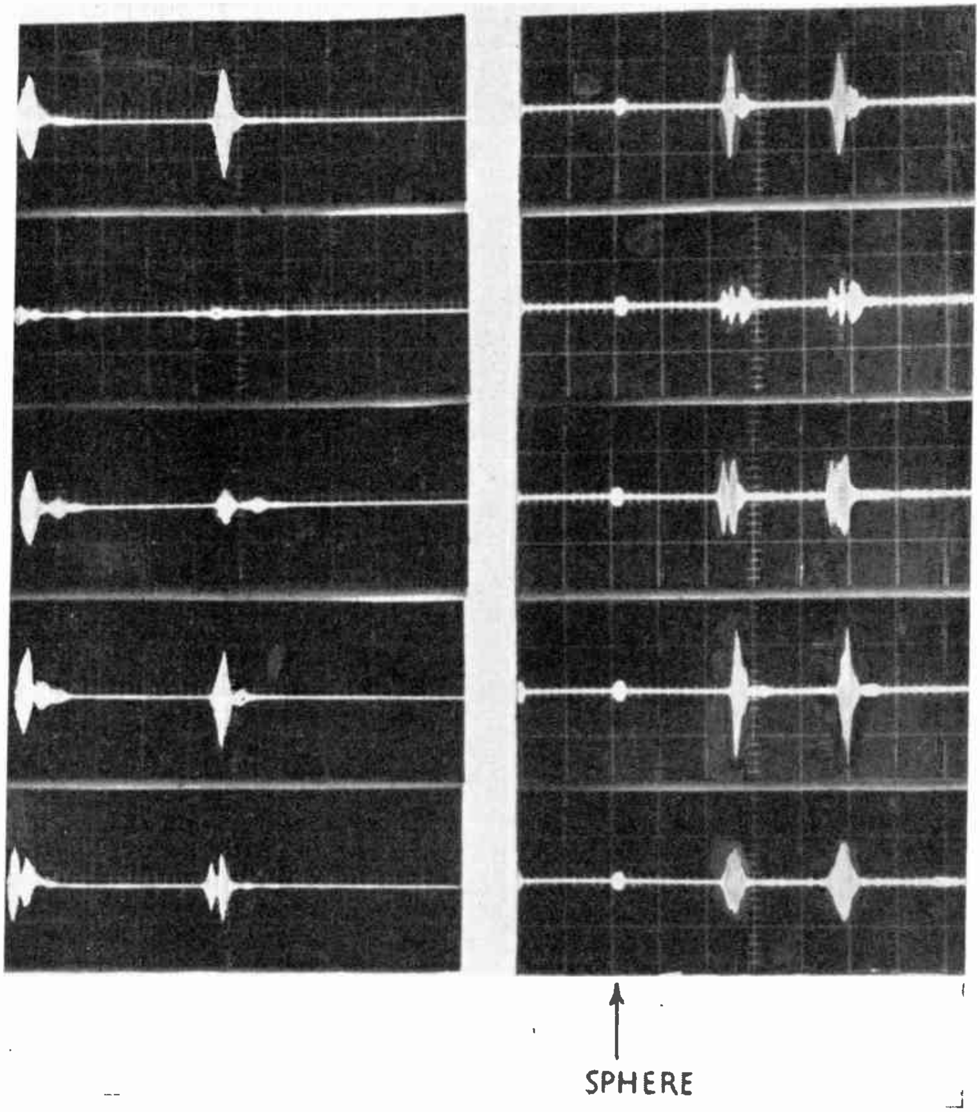
(b) the contours and composition of the particular sea-bed under investigation, and

(c) the time interval, after the inception of the echo pulse, at which the observation is made.

Changes in any of these will result in fluctuations of the echo pulse amplitude but only the first two are of any real interest and these will now be discussed in detail.

2.1. *Changes in the Geometrical Configuration*

The amplitude of an echo received from the sea-bed is obtained by summing, vectorially, at the receiving transducer all the signals reflected by the large number of scatterers on the bottom illuminated by the transmitter. These signals will arrive at the receiver with different phases and any change in the geometrical configuration, resulting from a combination of the vessel's three residual motions, will change the phase relationship between them, thereby changing the observed echo amplitude.



(*a*) Consecutive echo-sounder traces.

(*b*) Fluctuations of echoes reflected from 18 in diameter hollow spherical target one fathom from the bottom.

Fig. 1. Fluctuations in amplitude and shape of the carrier envelope.

To facilitate analysis of this process, these residual motions may be resolved into horizontal, vertical and angular rotational components and to assess their relative importance it is necessary to determine the displacement required to randomize successive echoes. This may be achieved using the following argument.

2.1.1. Horizontal displacement

Firstly consider the simple case of the transducers moving horizontally in relation to two discrete targets having uniform size, shape and composition. The resulting amplitude/displacement series has an auto-correlation function which takes the value zero when the displacement is sufficient to change the relative phases of the signals from the two targets by π radians. This means that the combined echo observed after the displacement will be completely unrelated to the initial echo.

Now, let these two discrete targets be replaced by two elemental areas of the insonified section of the sea-bed, normal to the direction of transducer displacement, such that each lies midway between the centre of the illuminated section and its extremities in the direction of displacement, one on either side of the centre. Once again signals from scatterers situated along these two elemental strips will be uncorrelated when the displacement is sufficient to change their relative phases by π. If the sea-bed is assumed to be composed of a number of discrete scatterers disposed with an approximately uniform density, it follows that the number under the transducers will be small, consequently even though the transducer's gain is large in this direction, the actual contribution to the observed echo received from the central area will be small. Furthermore, at the extremities of the illuminated section, where the number of scatterers is very much larger, the gain of the transducers approaches zero, so, once again, the contribution to the received echo will be small. By far the largest contribution must, therefore, come from the areas of the insonified sea-bed lying close to the positions corresponding to the 3 dB (half-power) directions of the transducers and it is in these critical areas that the two elemental strips are located.

It is therefore possible to estimate the horizontal and angular movements necessary to randomize the echoes from the whole illuminated section because this must be of the same order of magnitude as the displacements required to uncorrelate the echoes from both the elemental areas and the discrete targets.

From the geometrical configuration shown in Fig. 2(*a*) in which the vessel moves horizontally by an amount d_1 from the position when the signals from both elemental areas were in phase, it can be seen that the signal path length difference is

$$2\left\{\left[R^2+(r+d_1)^2\right]^{\frac{1}{2}} - \left[R^2+(r-d_1)^2\right]^{\frac{1}{2}}\right\}$$

which reduces to $4rd_1/R$ when $R \gg d_1$. Complete cancellation of the signals reflected by the two elements will occur when this path difference is equal to $\lambda/2$, therefore, the horizontal displacement necessary to produce completely uncorrelated echoes from the whole insonified section is given by

$$d_1 = \frac{R\lambda}{8r} \qquad \text{......(1)}$$

where $4r$ is the length of the illuminated section in the direction of travel.

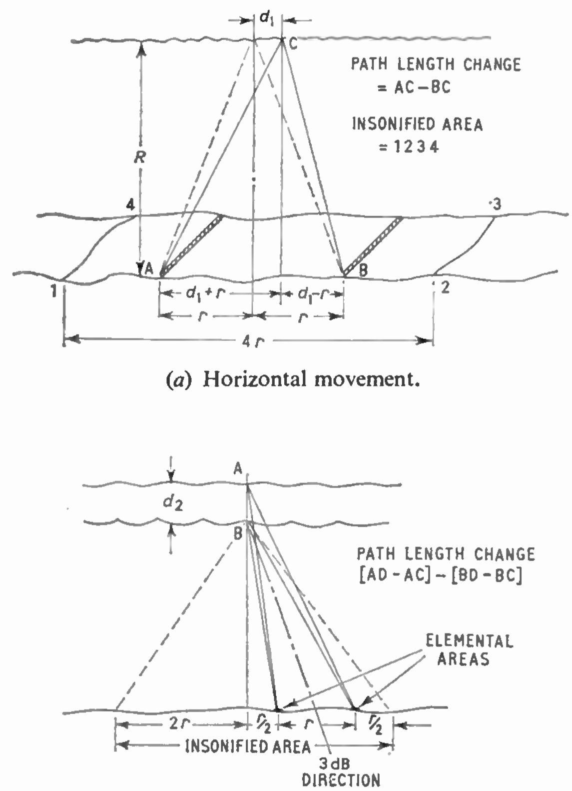
2.1.2. Angular displacement

Angular displacement deflects the acoustic beams from the vertical and causes the illuminated section to move across the bottom so giving rise to echo randomization in a manner very similar to that produced by horizontal motion. The roll angle necessary to produce a displacement equivalent to d_1 and hence to uncorrelate the bottom echo pulses is given by

$$\psi \simeq \tan^{-1}\frac{d_1}{R} \simeq \tan^{-1}\frac{\lambda}{8r} \qquad \text{......(2)}$$

2.1.3. Vertical displacement

When the vessel moves vertically, the relative phases of signals returned from scatterers situated at equal distances from the centre of the insonified section remain unchanged, consequently, the randomizing mechanism must be sought elsewhere. Two processes occur during the vertical motion; firstly, extra scatterers are insonified by the extremities of the sounding beam, and, secondly, the relative path lengths to scatterers which do not lie at equal distances from the centre change by small amounts.



(*a*) Horizontal movement.

(*b*) Vertical movement.

Fig. 2. Geometrical configurations for vessel movement.

If the vessel moves vertically through a distance d_2, the initial depth being R, the percentage increase in the number of scatterers insonified can be shown to be approximately equal to $(2d_2/R)/100$. In a practical case d_2/R may be as much as 5%, therefore the number of scatterers will be increased by 10%, but it must be remembered that these scatterers lie in a region of almost zero transducer gain, consequently their contribution to the fluctuations must be small.

It has already been argued that the most important component signal contributions come from those scatterers lying close to the 3 dB directions, therefore, the amplitude fluctuations must result from changes in the relative phases of signals returned from the scatterers positioned in elemental areas on either side of this 3 dB direction. The motion-induced signal path-length differences increase progressively as the elemental areas are separated and reach a maximum when one lies immediately under the vessel and the other lies at the extremities of the insonified section. By considering the product of transducer gain and the number of scatterers in any particular direction, it can be shown that the maximum permissible separation of the two elemental areas occurs when they lie midway between the vertical and the 3 dB directions, and the 3 dB directions and the limits of the insonified area, respectively. The geometrical configuration of this is shown in Fig. 2(*b*). The path-length difference between the two slant ranges changes, when the vessel moves vertically through a distance d_2, by an amount equal to twice the expression

$$\left\{\left[R^2+\left(\frac{3r}{2}\right)^2\right]^{\frac{1}{2}}-\left[R^2+\left(\frac{r}{2}\right)^2\right]^{\frac{1}{2}}\right\}-$$
$$-\left\{\left[(R+d_2)^2+\left(\frac{3r}{2}\right)^2\right]^{\frac{1}{2}}-\left[(R+d_2)^2+\left(\frac{r}{2}\right)^2\right]^{\frac{1}{2}}\right\}$$

In practice $R \gg d_2$ and the above expression reduces to $2r^2d_2/R^2$. The vertical displacement necessary to uncorrelate the bottom echoes is therefore given by,

$$d_2=\frac{\lambda R^2}{4r^2} \qquad \ldots\ldots(3)$$

The ratio of the horizontal to the vertical randomizing displacement is

$$\frac{d_1}{d_2}=\frac{r}{2R} \qquad \ldots\ldots(4)$$

and this will always be considerably less than unity in practice, which shows that the horizontal motions resulting from rolling and yawing are more important to the production of amplitude fluctuations than the vertical motion caused by pitching.

2.2. *Statistical Analysis of the Amplitude Fluctuations*

Further properties of the amplitude/displacement process can be determined using statistical methods. It is shown in the Appendix that the echo amplitude/displacement series obtained by moving the transducers progressively through distances which are small in comparison with the dimensions of the insonified area have the same statistical properties as the series obtained when the incremental displacement is larger than the illuminated section. (In practice, the former series is obtained if the echo-sounder is used when the vessel is at anchor, whilst the latter is obtained whenever the ship is steaming at a speed sufficient to change the illuminated area between successive transmissions.) The process can, therefore, be regarded as 'ergodic' which means that the statistical properties of all possible series of data relating to transducer movements in any direction over any contour configuration, assessed using any time interval between transmissions, can be characterized by the properties of a single time or displacement series of data obtained by moving the transducers progressively across one particular bottom configuration only, provided the bottom composition remains unchanged throughout. It is also shown in the Appendix that the echo amplitudes are distributed according to the Rayleigh probability density function

$$P(Z)=\frac{Z^2}{\sigma^2}\exp\left(-\frac{z^2}{2\sigma^2}\right)$$

2.3. *Changes in the Composition of the Sea-bed*

The size of the individual scatterers comprising the sea-bed under investigation will further affect the mean echo amplitude and its rate of change for a given transducer displacement. When the sea-bed is composed of a small number of approximately spherical scatterers whose diameter is large in comparison with the acoustic carrier wavelength, each will act as an isotopic radiator with an effective reflecting area corresponding to the diameter of its first Fresnel zone, i.e. $\pi\lambda(D-\lambda)$. The total energy reflecting area will then be only a fraction of the insonified area, consequently the mean echo amplitude will be small and the horizontal and vertical randomizing displacements will be independent of the scatterer size.

As the ratio of particle size to acoustic carrier wavelength decreases below unity, the bottom acts less like a diffuse scatterer and more like a perfect reflector. The number of scatterers contributing to each echo will be increased, which gives rise to a corresponding rise in the mean echo amplitude. When $D \ll \lambda$ specular reflections result which greatly reduce the effective size of the insonified area, thereby reducing the value of r in equations (1), (2) and (3) which means that echo randomization can only be accomplished by very much larger displacements of the transducers.

Under conditions of specular reflections, the relative phase differences between the various component signals becomes exceedingly small, so in spite of the reduction in the reflecting area, the mean echo amplitude increases as the scatterer size decreases. When $\lambda > 50D$, the effect of particle size can be neglected altogether and the amplitude fluctuations can be attributed to undulations of the bottom.

If the sea-bed is composed of scatterers varying in size, some larger and some smaller than the acoustic wavelength, the mean signal level will be decreased and its rate of change with transducer displacement will be increased as the number of large scatterers is increased. These effects are due to the resulting effective reduction in the number of reflecting particles and greater disparity between the phases of the various signal components.

3. Experimental Investigation

3.1. *The Scale Model*

To enable a thorough experimental investigation to be carried out it is imperative to operate the echo-sounding system in an operational environment which provides the following pre-requisites:

(1) complete freedom from thermal and other inhomogeneities within the medium, turbulence, etc.,

(2) facilities for obtaining different contours upon a bottom of known fixed composition,

(3) facilities for changing the composition of the bottom and measuring the subsequent spread of scatterer sizes, together with,

(4) accurately controlled and measurable, horizontal and vertical displacements of the transducers over the bottom from a fixed datum.

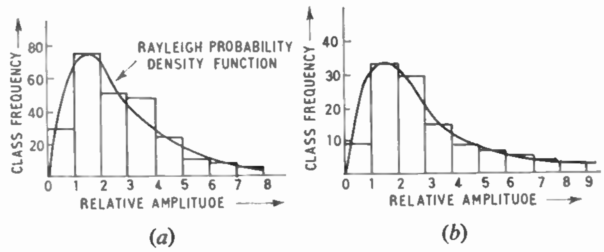
All these conditions were adequately satisfied by the use of a scale model in a water tank measuring 3 ft × 3 ft × 2 ft deep. Two 517 kc/s lead zirconate elements, one to transmit, the other to receive, and each measuring 12 mm × 6 mm, were mounted on a carriage moving along a rail bolted rigidly to the top of the tank. This permitted a controlled horizontal movement of over 50 cm and a vertical motion of 5 cm with positional accuracies of 0·02 mm and 0·05 mm respectively.

Owing to the small dimensions of the tank, acoustic energy transmitted and received by the side-lobes of the transducers was reflected from the side walls and appeared on the display very shortly after the required bottom echo. To avoid the consequences of this, a pulse length of only 0·1 ms was used throughout and all amplitude measurements were taken at a time $t = 0{\cdot}04$ ms after the pulses inception.

At the high acoustic carrier frequency used, penetration of the bottom material by the acoustic energy was very small and a minimum bottom thickness of only ½ in was used, on top of which contours with a maximum height of ½ in were applied. The materials used to simulate the sea-bed, plaster of paris, sand, gravel and pebbles, were chosen to provide a wide range of scatterer sizes from $\lambda/100$ to 12λ and the problem of obtaining a 'statistically random' bottom was solved in one of two ways depending upon the size of the particles constituting the bottom. For the small particle sizes irregular, non-cyclic agitation of the material was used and this caused it, after settling, to be disposed in a random manner. The larger particles were scattered uniformally upon the surface of the water, their uneven shape subsequently caused each one to follow a different irregular path to the bottom, thereby building up a randomly distributed layer.

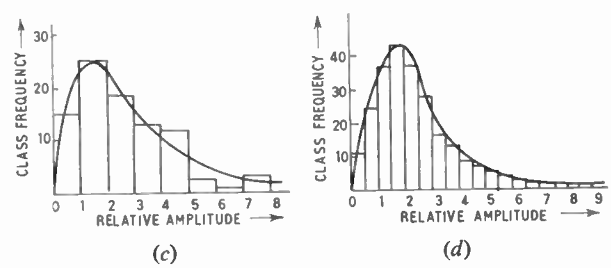
3.2. *The Verification of the Ergodic Nature of the Fluctuations*

A mixture of silver sand and gravel, in the ratio of 4 to 1 by weight, was used to simulate the sea-bed, the particle sizes being in the range $1/12\lambda$ to 3λ. The echo pulse amplitude, measured as the open circuit voltage of the receiving transducer, was recorded as the transducers were moved horizontally in 240 1 mm increments and this process was repeated ten times using a different contour configuration each time. The mean amplitude of the ten sets of data was 268 mV, the maximum deviation from the mean being only 2%. The amplitude distribution of the 10 series is shown in Fig. 3(*a*) and is in close agreement with the theoretical Rayleigh distribution function and that derived from



(*a*) Horizontal increments simulated.
Acoustic frequency = 517 kc/s. Pulse length = 0·1 ms.
Bottom composition—sand and gravel.

(*b*) Small horizontal increments actual, vessel anchored.
Acoustic frequency = 37 kc/s. Pulse length = 0·25 ms.
Bottom composition—sand, mud and debris.



(*c*) Large horizontal increments simulated.
Acoustic frequency = 517 kc/s. Pulse length = 0·1 ms.
Bottom composition—sand and gravel.

(*d*) Large horizontal increments actual, vessel moving at 4 knots.
Acoustic frequency = 37 kc/s. Pulse length = 0·25 ms.
Depth of sea-bed = 14 fathoms ± ½ fathom.

Fig. 3. Amplitude/displacement series.

data recorded whilst the R.R.S. *Discovery II* was at anchor off Falmouth (Fig. 3(*b*)).

Echo pulse amplitudes were also recorded as the transducers were moved in 10 1 cm increments from a fixed datum over 90 different contour configurations. The mean amplitude of the ten series was 256 mV, the maximum deviation from this being 5%. The corresponding amplitude distribution is shown in Fig. 3(*c*). Figure 3(*d*) shows the echo amplitude distribution for the analogous case when the R.R.S. *Discovery II* was steaming at a speed sufficient to change the insonified area between successive transmission.

The close agreement between the amplitude distribution and the mean amplitudes (together with further statistical properties not mentioned here) shows conclusively that the echo amplitude/displacement process is indeed ergodic.

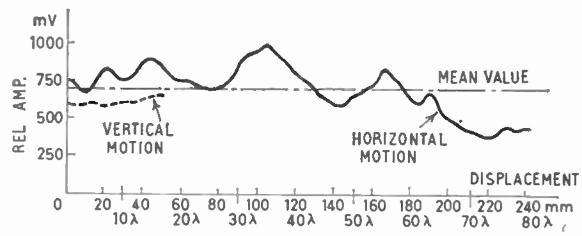
3.3. *The Effect of Scatterer Size*

Six different materials were used to form the bottom and these are listed with their corresponding spread of particle sizes:

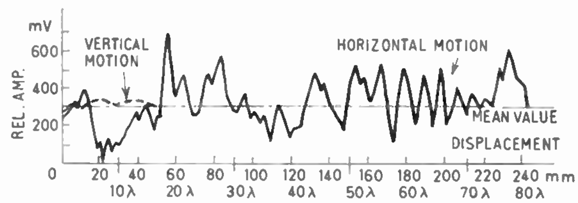
(1) Plaster of paris	$\lambda/100$–$\lambda/30$
(2) Silver sand	$\lambda/12$–$\lambda/3$
(3) Silver sand and gravel (4 to 1)	$\lambda/12$–3λ
(4) Silver sand and pebbles (4 to 1)	$\lambda/12$–12λ
(5) Gravel	λ–3λ
(6) Pebbles	4λ–12λ

It was impossible to ensure that the contours on the bottom were the same in each case, but this is of no consequence owing to the ergodic nature of the process.

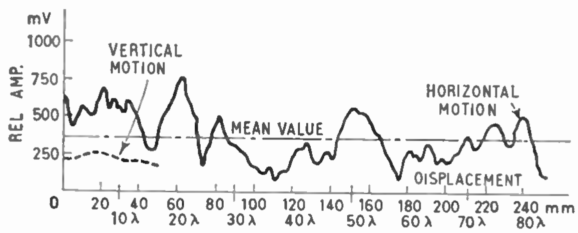
Echo amplitude was recorded as the transducers were moved horizontally in 240 1 mm increments and vertically in 50 1 mm increments (each motion being made independently) over the six different bottoms in turn. Typical graphs showing the dependence of the echo amplitudes upon movement as a function of bottom composition are given in Fig. 4 and relevant information extracted from them is summarized in Table 1.



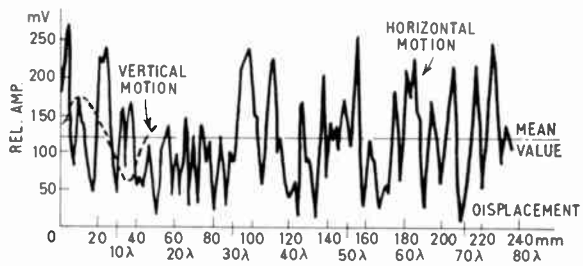
(*a*) Nature of bottom—plaster of paris. Grain size $\lambda/100$ to $\lambda/30$.



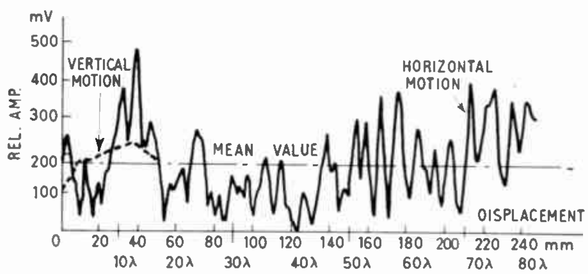
(*d*) Nature of bottom—sand and pebbles. Scatterer size $\lambda/12$ to 12λ.



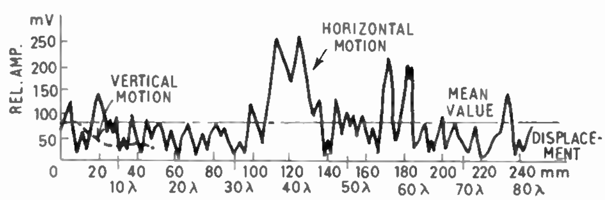
(*b*) Nature of bottom—sand. Grain size $\lambda/12$ to $\lambda/3$.



(*e*) Nature of bottom—gravel. Scatterer size λ to 3λ.



(*c*) Nature of bottom—sand and gravel. Grain size $\lambda/12$ to 3λ.



(*f*) Nature of bottom—pebbles. Scatterer size 4λ to 12λ.

Fig. 4. Amplitude fluctuations due to the motion of the transducers. Acoustic carrier frequency = 517 kc/s. Transmitted pulse length = 0·1 ms. Amplitude at echo recorded after 0·04 ms. Contours $\leqslant \frac{1}{2}$ in for (*b*), (*c*), (*d*) and (*e*).

Table 1

Material	Scatter size	Mean signal level $\frac{1}{n}\sum_{i=1}^{n} A_i = \bar{A}$	Mean deviation $\frac{1}{n}\sum_{i=1}^{n} A_i - \bar{A}$	Randomizing horizontal	Displacement vertical (approx.)
Plaster of paris	λ/100–λ/30	705 mV	13%	30 mm	—
Sand	λ/12–λ/3	362 mV	37%	10 mm	—
Sand and gravel	λ/12–3λ	180 mV	46%	4 mm	—
Sand and pebbles	λ/12–12λ	308 mV	40%	1 mm	15 mm
Gravel	λ–3λ	118 mV	48%	1 mm	15 mm
Pebbles	4λ–12λ	74 mV	52%	1 mm	15 mm

This shows that the mean signal level decreases and the fluctuation rate for a given displacement increases as the scatter size is increased. One important conclusion that can immediately be drawn from this particular experiment is that there must always be some scatterers, larger than the acoustic wavelength present on the sea-bed if echo randomization by small movement is to take place.

The theoretical randomizing horizontal and vertical displacements calculated using eqns (1) and (3) and the parameters of the scale model for which $R = 35$ cm, $\lambda = 2{\cdot}91$ mm and $4r = 41$ cm are

(a) horizontal displacement $d_1 = 1{\cdot}22$ mm

(b) vertical displacement $d_2 = 8{\cdot}8$ mm.

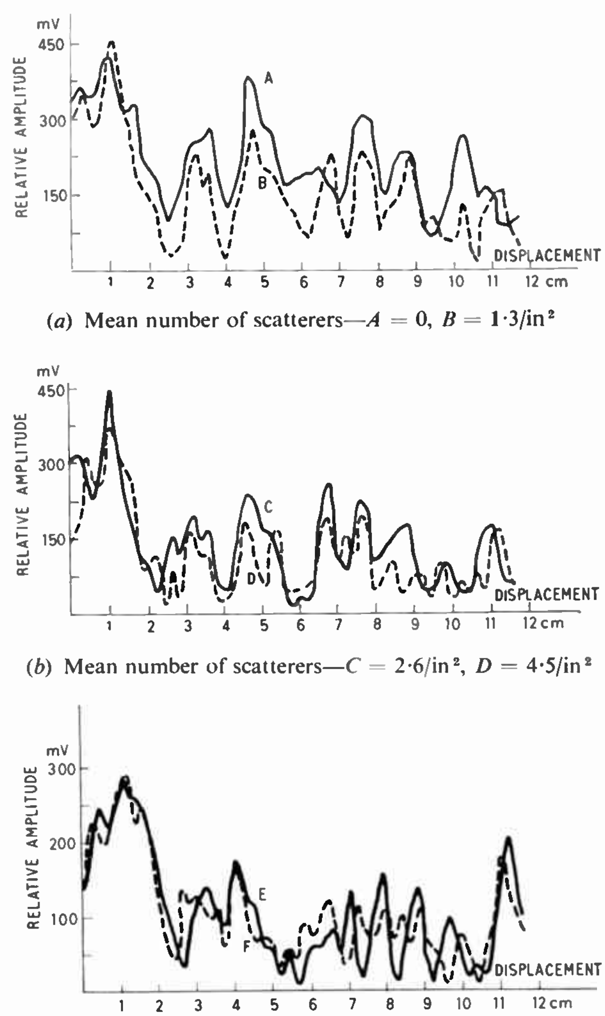
A comparison with the values observed, when scatterers larger than the acoustic wavelength are present on the bottom (Table 1), shows that the horizontal displacements are in agreement but the vertical displacements are less so, presumably due to the effect of the change in the number of scatterers insonified.

Figure 5 was derived from data obtained by moving the transducers horizontally in 120 2 mm increments over six different bottoms each having the same contour but a different composition. This was achieved by setting up suitable contours in silver sand and progressively adding more and more scatterers, in the form of gravel, in a random manner between each set of measurements. This demonstrates that as the concentration of large scatterers is increased so the mean signal is decreased and the rate of change of amplitude is increased.

4. Conclusions

These can be summarized as follows:

(1) The amplitude of the echo reflected by the sea-bed fluctuates due to the three residual motions of the vessel, pitching, rolling and yawing, the amount of



(a) Mean number of scatterers—$A = 0$, $B = 1{\cdot}3/\text{in}^2$

(b) Mean number of scatterers—$C = 2{\cdot}6/\text{in}^2$, $D = 4{\cdot}5/\text{in}^2$

(c) Mean number of scatterers—$E = 9{\cdot}8/\text{in}^2$, $F = 15/\text{in}^2$

Fig. 5. Amplitude fluctuations. Effect of bottom composition. Transmitted pulse duration = 0·1 ms. Acoustic frequency = 517 kc/s. Amplitude measured after 0·04 ms. Nature of bottom —sand and gravel. Constant contours λ/12 to 3λ.

movement required to randomize the echoes being very small, and approximately equal to:

(a) pitching $= \dfrac{\lambda R^2}{4r^2}$

(b) rolling $= \tan^{-1} \dfrac{\lambda}{8r}$

(c) yawing $= \dfrac{R\lambda}{8r}$

(2) 'Rolling' and 'yawing' are far more important to the production of amplitude fluctuations than 'pitching'.

(3) Statistical properties of a series of data, relating to movement in any direction over any contour configuration assessed using any time interval between transmissions, can be characterized by the properties of a time or displacement series of data obtained by moving the point of observation progressively across one particular bottom only, provided the bottom composition remains unchanged throughout.

(4) Some scatterers larger than the wavelength must always be present on the sea-bed if small movement randomization is to take place.

5. References

1. L. Liebermann, "Thermal inhomogeneities in the ocean", *J. Acoust. Soc. Amer.*, **23**, p. 563, 1951.
2. D. C. Whitmarsh, E. Skudrzyk and R. J. Urick, "Forward scattering of sound in the sea and its correlation with the thermal micro-structure", *J. Acoust. Soc. Amer.*, **29**, p. 1123, 1957.
3. P. G. Bergmann, "Propagation of radiation in a medium with random inhomogeneities", *Phys. Rev.*, **70**, p. 486, 1946.
4. D. Mintzer, "Wave propagation in a randomly inhomogeneous medium, I and II", *J. Acoust. Soc. Amer.*, **25**, p. 922 and p. 1107, 1953.
5. J. S. Bendat, "Principles and Applications of Random Noise Theory", Chap. 3, p. 80. (Chapman & Hall, London, 1958.)
6. Ref. 5, Chap. 3, p. 135.

6. Appendix: Statistical Properties of the Amplitude Fluctuations

6.1. *Small Horizontal Displacements of the Transducers*

Consider the series of echo-pulse amplitudes, measured throughout at fixed time after the inception of the echo pulse reflected from a particular sea-bed represented by the index k_0, extracted from an infinity of possible sea-bed configurations $k_0, k_1, k_2 \ldots k_n$. Displacement of the transmitting and receiving transducers in any direction, e.g. the x direction of Fig. 6, leads to the amplitude displacement function, $Z_{k_0}(x)$. At a particular position $x = x_0$, the echo amplitude is obtained from a vectorial summation, carried out at

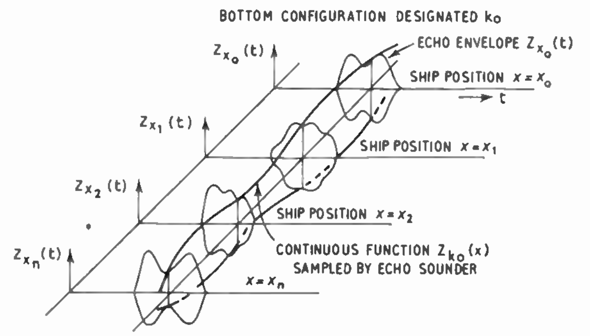


Fig. 6. Three-dimensional representation of generation of the amplitude displacement function $Z_{k0}(x)$.

the receiving transducer, of the signals reflected from a large number of scatterers q disposed in a random manner. Thus:

$$Z_{k_0}(x_0) = \sum_{n=1}^{q} A_n \cos(\omega_c t + \phi_n)$$

If now the transducers are moved to the position $x = x_1$ such that the total horizontal displacement $x_1 - x_0$ is very much less than the dimension of the insonified section of the sea-bed in the direction of the displacement, most of the scatterers will be common to both positions and

$$Z_{k_0}(x_1) = \sum_{n=r}^{q+r} B_n \cos(\omega_c t + \theta_n)$$

where r is the number of scatterers added to the leading edge and subtracted from the trailing edge of the insonified section assuming the scatterers to be distributed with an approximately uniform density. If it is further assumed that the scatterers are of approximately uniform size, shape and composition

$$\sum_{n=1}^{r} A_n \simeq \sum_{n=r}^{q+r} B_n$$

and the echo amplitude change will result mainly from changes in the relative phases between the signals reflected by scatterers common to both positions.

Although ϕ_n and θ_n are related, the actual relationships can be calculated for only the simplest of bottom contour configurations, consequently it is impossible to assess the echo amplitude change

$$Z_{k_0}(x_0) - Z_{k_0}(x_1)$$

for any general case. This renders the derivation of a general expression for the 'continuous' amplitude/displacement function (obtained as x takes all values from x_0 to x_n progressively) impossible also. 'Continuous' is used here in the sense that the ideal function possesses no discontinuities; however, they do occur in practice from the use of pulsed carrier transmissions which sample the function. It can be argued however, that theoretically the function may take any value between zero and infinity, therefore it can be represented in Fourier Series form 5:

$$Z_{k_0}(x) = \sum_1^\infty a_m \cos m\omega t + \sum_1^\infty b_m \sin m\omega t$$

where a_m and b_m can take any value between $-\infty$ and $+\infty$ and are mutually independent, random variables, possessing equal auto-correlation products and zero cross-correlation product.

For every other sea-bed configuration it is possible to write a similar expression:

$$\left.\begin{aligned} Z_{k_1}(x) &= \sum_1^\infty a_{m_1} \cos m\omega t + \sum_1^\infty b_{m_1} \sin m\omega t \\ Z_{k_2}(x) &= \sum_1^\infty a_{m_2} \cos m\omega t + \sum_1^\infty b_{m_2} \sin m\omega t \\ Z_{k_n}(x) &= \sum_1^\infty a_{m_n} \cos m\omega t + \sum_1^\infty b_{m_n} \sin m\omega t \end{aligned}\right\} \quad \ldots\ldots(5)$$

If the composition of the sea-bed is unchanged throughout, this ensemble of functions must possess identical statistical properties; thus data properties relating to transducer movements over all possible sea-bed contour configurations will be completely characterized by the properties of data obtained using any one typical configuration, i.e.

$$\tilde{Z}_{k0}(x) = \tilde{Z}_{k1}(x) = \tilde{Z}_{kn}(x)$$

(The superscript $\sim$ denotes some statistical property of the functions.)

6.2. *Large Horizontal Displacement of the Transducers*

If the amplitude of the bottom echo is measured with the transducer in a fixed position, say at $x = x_0$, a series of varying echo amplitudes can be obtained

$$Z_{x0}(k)$$

as k is varied, i.e. as the contour configuration is changed. (This corresponds, in the practical case of the echo-sounder, to a transducer displacement greater than the dimensions of the insonified section taking place during the interval between successive transmission.) As before, this amplitude/displacement function may assume any value between zero and infinity and it can also be represented in Fourier Series form:

$$Z_{x_0}(k) = \sum_1^\infty A_p \cos p\omega t + \sum_1^\infty B_p \sin p\omega t$$

A_p and B_p are independent, random variables.

Further amplitude series can be obtained for each of a large number of discrete transducer positions, $x = x_1, x_2 \ldots x_n$ giving the functions:

$$\left.\begin{aligned} Z_{x1}(k) &= \sum_1^\infty A_{p_1} \cos p\omega t + \sum_1^\infty B_{p_1} \sin p\omega t \\ Z_{x2}(k) &= \sum_1^\infty A_{p_2} \cos p\omega t + \sum_1^\infty B_{p_2} \sin p\omega t \\ Z_{xn}(k) &= \sum_1^\infty A_{p_n} \cos p\omega t + \sum_1^\infty B_{p_n} \sin p\omega t \end{aligned}\right\} \quad \ldots\ldots(6)$$

Once again, provided the bottom composition remains constant, each member of this ensemble of functions will possess identical statistical properties, i.e.

$$\tilde{Z}_{x0}(k) = \tilde{Z}_{xn}(k)$$

The amplitude/displacement functions constituting ensembles (5) and (6) all have the same form and therefore they will all have the same probability density function which can readily be shown[6] to be the Rayleigh probability density function

$$P(Z) = \frac{z}{\sigma^2} \exp\left(\frac{-Z^2}{2\sigma^2}\right)$$

If the composition of the sea-bed is the same for both ensembles (5) and (6) it is apparent that the statistical properties of the amplitude fluctuations, resulting from (a) motion of the transducer over one particular contour configuration, and (b) randomization of the contours completely between successive transmissions, will be identical, i.e.

$$\tilde{Z}_{kn}(x) = \tilde{Z}_{xn}(k)$$

and the whole process can be regarded ergodic (similar arguments can be used to obtain the same result in the case of angular and vertical motions).

Manuscript first received by the Institution on 21st June 1962 and in final form on 6th February 1963. (Paper No. 836/SS6.)

POINTS FROM THE DISCUSSION

Mr. R. J. Urick†: The figure used to illustrate fluctuation showed echoes of a sphere as well as the bottom return. These sphere echoes could be used to calibrate the system and the bottom echoes could then be used to determine the scattering coefficient of the bottom. In fact, the ratio of amplitudes of sphere echo and bottom echo is a direct measure of bottom scattering strength at normal incidence, when the size of the insonified area is known. Values of this quantity would be most useful if they could be obtained and published.

The author (*in reply*): The bottom scattering coefficient can be obtained from the data of Fig. 1(*b*) provided the

† U.S. Naval Ordnance Laboratory.

observed target sphere amplitude is corrected to take into account the difference between the depths of the sphere and the sea-bed. Furthermore, because the sphere was situated in the direction of maximum gain of the transducer whereas the bottom echo results from signals reflected by scatterers lying in directions of varying transducer gain, the size of the insonified area must also be corrected to obtain an effective area over which the gain has a constant value equal to its maximum. Assuming the transducers to possess piston-like properties the effective 'constant gain' beamwidth is given by the expression $2 \sin^{-1} 0{\cdot}59\ \lambda/l$ for a rectangular transducer, where λ is the acoustic wavelength and l is the length of the transducer face.

Applying these considerations to the practical sounding system, which operated at a signal frequency of 37 kc/s using a transducer with an active face measuring 13·5 cm and 6 cm, gives the effective size of the insonified area, at the working depth of 14 fathoms, as 30 ft × 73 ft.

Owing to the complex structure of the bottom echo, its mean amplitude must be computed before it is compared with the sphere echo amplitude. The ratio of the sphere echo amplitude to the mean bottom echo amplitude, computed using data obtained from Fig. 1(*b*), lies in the range 1 : 1·3 to 1 : 3·8 and after applying the depth correction these are modified to 1 : 1·8 to 1 : 5·3.

Assuming a reflection coefficient of 0·95 for the air-filled, cast-iron sphere, these ratios yield sea-bed scattering coefficients lying in the range $2{\cdot}7 \times 10^{-3}$ to $2{\cdot}4 \times 10^{-2}$. This range of coefficients applies to a sea-bed composed largely of a mixture of sand and mud plus a little river debris but their actual proportions and their disposition is, unfortunately, unknown.

Dr. R. W. G. Haslett†: Some work conducted at 30 kc/s using a vertical sound beam in shallow water‡ has indicated that when the beam is wide and the sea-bed is smooth, the variation in the strength of the sea-bed echo has a small mean deviation (11·7% of the mean value) and that the variations may be attributed entirely to the motion of the vessel. Also, it appeared that the effective area of the sea-bed was contained within a small angle.

This contrasts with Fig. 3 of the present paper in which the amplitudes vary more widely and, in fact, sometimes fall to zero. Can the theory put forward by Dr. Gazey be extended to include the above case?

The author (*in reply*): Dr. Haslett's work, using a vertical sound beam over a smooth, sandy sea-bed, is an example of the phenomena discussed in Section 2.4. Here it is argued that as the individual scatterer size becomes small compared with the wavelength of the acoustic carrier, so the bottom acts as a reflector and not as a diffuse scatterer. Under these conditions, the effective area of the sea-bed will be contained within a small angle, a fact observed by Dr. Haslett, and the fluctuations may then be attributed either to the rolling and pitching motions of the ship or to small undulations of the sea-bed, if present.

Figure 4(*a*) shows an example of the fluctuations obtained from a slowly undulating (simulated) sea-bed when the scatter size to acoustic wavelength ratio is considerably less than unity and is very similar to the results obtained by Dr. Haslett under similar conditions. Figure 3 contrasts with Dr. Haslett's observations and the fluctuations of Fig. 4(*a*) because although the same scatter size acoustic wavelength ratio applied, the sea-bed was no longer smooth or slowly undulating being subjected to the scouring action of the effluent from the River Fal, which would give rise to small, closely packed undulations, each of which could be regarded as a scatterer larger than the acoustic wavelength thereby giving rise to the large echo amplitude fluctuations recorded.

† Kelvin Hughes Division of S. Smith & Sons (England).

‡ R. W. G. Haslett, *J. Brit.I.R.E.*, **22**, pp. 38–42 and Figs. 8 and 9, 1961.

U.D.C. 53.084: 621.395.625.3: 621.774

Linear Measurement of Steel Tubes in Motion by Combining Magnetic Imprinting with a Wheel Counter

By

R. ASHFORD†

AND

P. HUGGINS (*Member*)‡

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

Summary: A method of measuring the throughput of steel tube or bar from a cold rolling mill or on a conveyor is described. Magnetic spots are created at 1 ft intervals along the extensive length: these are counted, thereby recording the aggregate footage. A wheel of 1 ft circumference is also held in contact with the material and this is used to measure that fraction of a foot not measured by the foot counter. The combination of footprints and wheel makes for an accurate system of measurement at high speeds where the surface conditions of the steel are such that a simple wheel and counter would be highly inaccurate due to slip. Laboratory experiments are described showing the potential accuracy of the system. A proposed application for the measurement of steel tubes is described and its probable performance evaluated.

1. Introduction

In steel tube manufacture there is a production requirement for knowledge of the aggregate footage of a batch of tubes passing along a roller conveyor. In one such requirement the tubes were of random length (ranging from 8 to 60 ft) and spaced out so that there was a gap of at least 6 ft between tubes. They had uniform outside diameter in any one batch ($\frac{1}{4}$–$1\frac{1}{2}$ in). As is often the case the tube may be accelerating and decelerating during the measurement. A typical speed range is likely to be 100 ft/min to 300 ft/min. In this particular requirement the accuracy wanted was $\pm 1\%$ over a batch of not less than twenty tubes.

The obvious method of measurement, a friction wheel bearing on the tube surface which drives a counter calibrated in linear scale, is inappropriate because the tube is likely to be oily from a previous process. For every tube, such a wheel would have to start from rest, then pick up the tube speed. It has been found in practice that if the measuring wheel is adequately loaded so that it grips the leading end of the tube effectively, the wheel periphery soon becomes damaged and worn because the ends of the tubes are sharp. There is also danger of marking some tubes if the effective wheel pressure is adequate. Again, if a wheel were used it would be necessary to lower it at the instant the tube's leading edge had passed the wheel centre line. Apart from this difficulty there will probably be an appreciable slip due to the range of tube speed, variations in diameter and surface conditions; this would certainly exceed $\pm 1\%$.

† Tube Investments Ltd., Steel Tube Division, Development Engineering Department, Walsall, Staffs.

‡ Formerly with Tube Investments Ltd.; now with Girling Ltd., Birmingham, 11.

2. Magnetic Imprinting

A known method of measuring extensive length makes use of electro-magnetic imprinting.§ Two electro-magnetic heads, one for writing and one for reading, are located immediately above the steel and in line with the direction of travel (See Fig. 1). The heads are some arbitrary distance apart, and connected to an amplifier so that a spot appearing in the steel under the read head immediately triggers the write head, thus producing a further imprinted spot in the steel. These spots are counted when imprinted, and registered as the aggregate length in terms of the arbitrary distance between read and write heads.

Electromagnetic imprinting has two merits: (*a*) It does not depend on friction (being non-contacting), (*b*) it is relatively independent of feed speed variations.

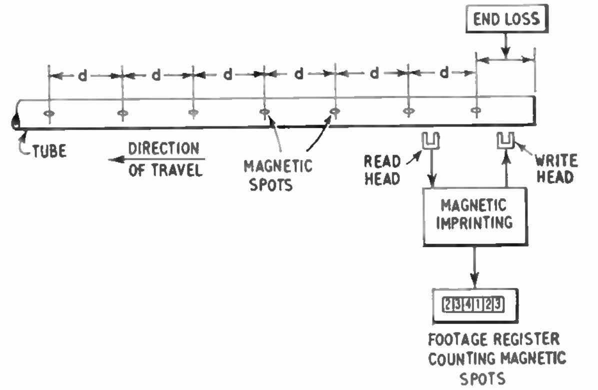


Fig. 1. Technique of length measurement by electromagnetic imprinting.

§ W. C. George, "Measuring strip steel without contact," *Instrum. Soc. Amer. J.*, No. 1, pp. 80–3, 1960.

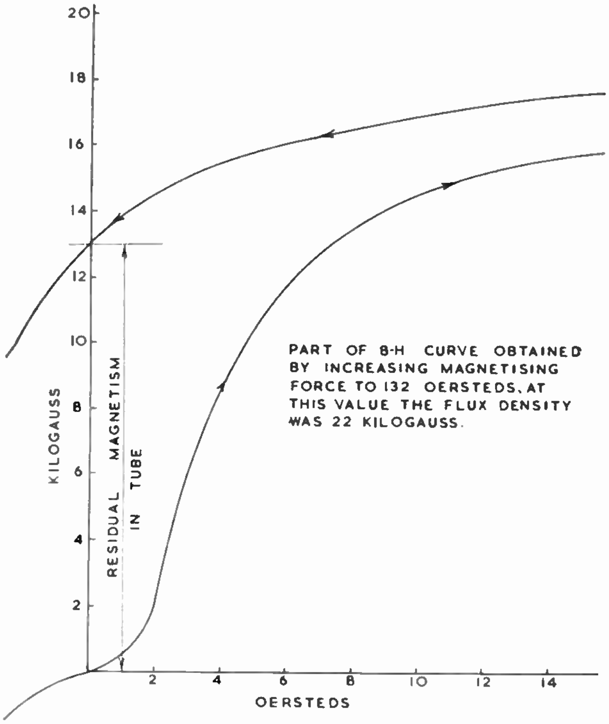


Fig. 2. Magnetic retentivity of typical refrigerator tube ⅞ in o.d. × 0·06 in wall.

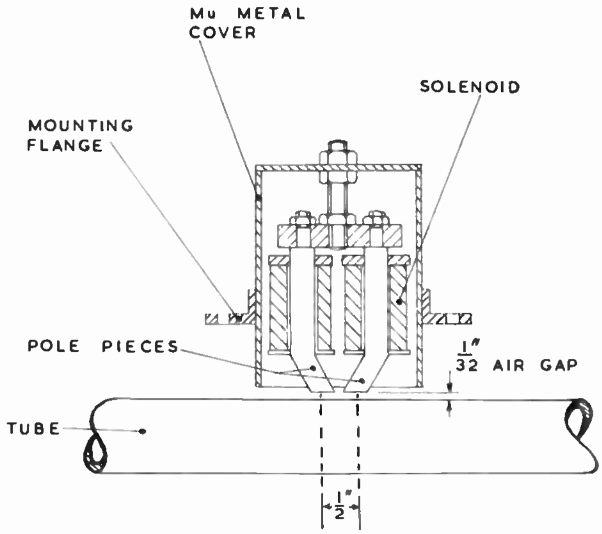


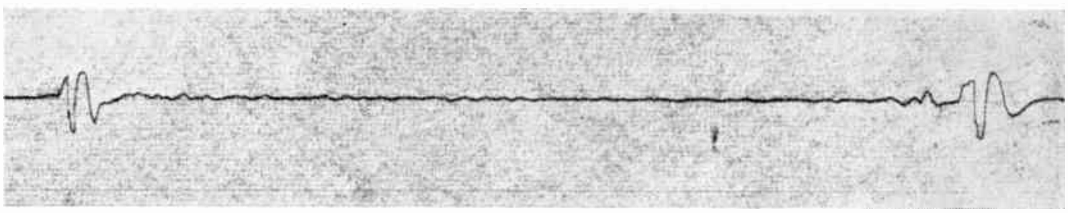
Fig. 3. Cross-section of a read and write head.

To establish feasibility, an experimental system was set up on a live conveyor in the laboratory. The results of the experiments indicated that it is unlikely that on batches of short tubes an accuracy of $\pm 1\%$ could be obtained by the electromagnetic imprinting alone. A greater accuracy can be obtained by including a friction wheel in the system, as described later.

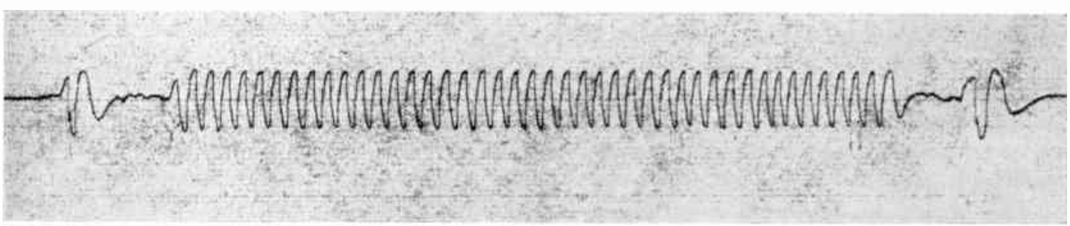
3. Experimental Work Relating to Magnetic Imprinting

Magnetic retentivity, reading/imprinting parameters, and tolerances of the system were studied in three sets of experiments.

In the first of these, samples of tube that would pass through the conveyor line were tested to obtain data on the magnetic properties of the steel. The results of an experiment to show the magnetic retentivity of a typical mild steel refrigerator tube is shown in Fig. 2. It will be seen that the density of the magnetic flux retained by the tube is 13 000 gauss: this is a readily detectable flux density. Magnetized tubes appeared to retain their magnetism in discrete areas for several days. It was observed that tubes had to be handled fairly gently if the magnetic spots



(a) An unmagnetized tube, 5 ft long, travelling at 100 ft/min.



(b) Same tube after magnetic imprinting at 1 in intervals throughout length. Note absence of spot signals within 9 in of each end.

Fig. 4. Trace of the recording head amplifier signal.

were to be retained. Experiments showed that if tubes were dropped from a height of 2 ft, no trace of the magnetic areas could be found.

In a second experiment to examine the system two electro-magnets were made (as shown in Fig. 3) and were mounted above a 1 in diameter tube resting on a vee roller conveyor with an air gap of 0·050 in between the pole face of the magnet and the tube. Magnetic spots were created by the write magnet. The tube was then traversed at 100 ft/min under the read head and the output was observed on an oscilloscope. The voltage induced in the read head by a magnetic spot is of the form shown in Fig. 4.

In the measuring system a magnetic spot passing under the read head was required to initiate the printing of a further magnetic spot. The read and write heads were electrically connected so that the amplified output from the read head caused the write head to place a magnetic mark in the tube. If the spots were placed closer together than 1 inch then the areas become less defined. It was found that at 1 inch separation there was interaction between the heads, which caused an uncontrolled state of oscillation. The heads had to be separated by at least 6 in to stop this. Shielding the heads with magnetic materials and muting the read head amplifier during the write pulse were tried without success. The presence of the tube enhanced the coupling between the heads. The ends of the tube induced large signals into the read head as shown in Fig. 4. It was noted that the imprinted spots vanished, or were variable, within 9 in of either end of the tubes (Fig. 4(*b*)).

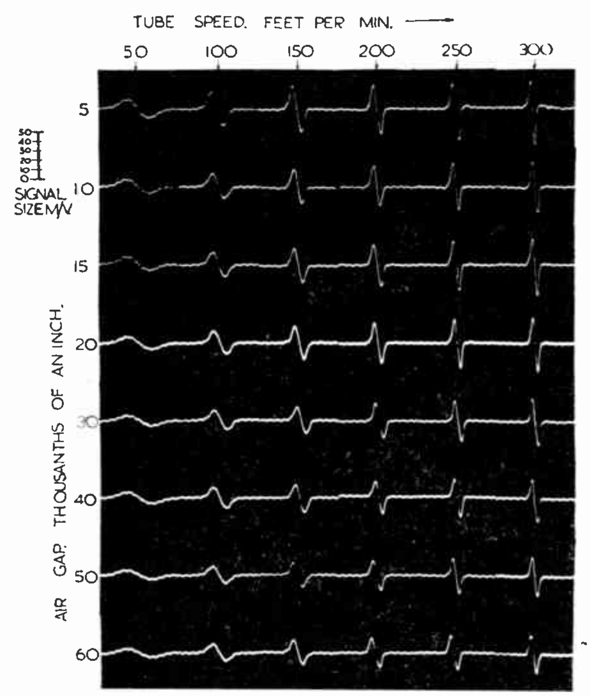


Fig. 5. Oscilloscope traces of signal amplitude variation vs speed and air-gap.

The third experiment was designed to study the effect of air-gap/speed changes on the output signal. A single magnetic spot was created in the tube which was then suspended on long supports and made to swing to and fro over a read magnet. The amplitude of swing was varied to give a predetermined velocity of magnetic spot travel over the read magnet. The air-gap was varied between 0·005 in and 0·060 in and a series of oscilloscope traces obtained (Fig. 5). From these the expected performance has been estimated.

4. General Comments on Experiments on Magnetic Imprinting

The experiments demonstrate the electromagnetic imprinting system to be feasible in the case of mild steel tubes. However the unpredictability of spot behaviour within 9 in of the tube ends and the instability of the systems when the heads are less than 6 in apart, make it desirable that the measuring increment should be (conveniently) 1 ft. In this event the fractional 'end loss' error, (see Fig. 1) could be 10% of the total length of one 8 ft tube, and is likely to be in excess of ±1% for small batches of such tubes even though an 'average' end loss of 6 in per tube is added to the measurement (see Appendix 1) by way of compensation.

5. Proposed System: Footprint and Wheel

To reduce the error due to end loss the system has to be extended. A friction wheel which adheres to the tube, and emits pulses, is used to measure the end piece. There are two reasons why this should be satisfactory despite our statement that friction wheel measurement alone is unacceptable (Section 1).

(*a*) The bulk of the tube will have passed under the measuring wheel; therefore the initial skidding period when the wheel tries to grip the tube is avoided.

(*b*) The inaccuracy of measurement applies to the increment at the tail end of the tube only. Expressed as a percentage of the whole tube lengths this percentage slip is reduced by at least one order of magnitude.

Figure 6 shows the arrangement of the hybrid system. The read/write heads are 1 ft apart, and the measuring wheel is 1 ft in circumference. The measuring wheel contains a photo-electric pulse generator emitting 16 pulses per revolution. Three limit switches (photo-electric) spaced as indicated (Fig. 6) control the measuring sequence.

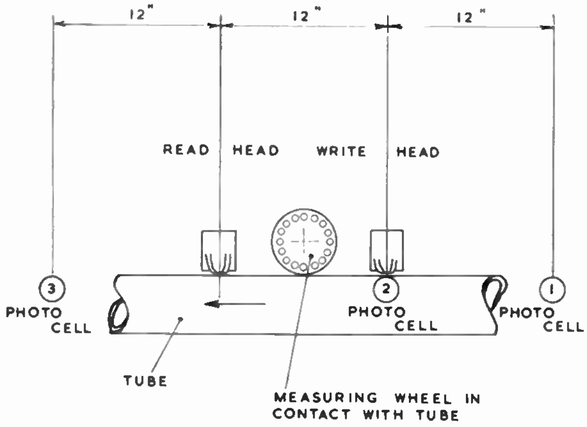


Fig. 6. Hybrid system showing the location of the limit switches.

In order to combine the two measurement techniques in this way, the following requirements must be satisfied.

(*a*) The apparatus must be muted until the first spot is imprinted: in the present system this is 2 ft (i.e. distance between PC2 and PC3) from the leading end of the tube.

(*b*) Magnetic imprinting should then continue as previously described (Section 2). The number of imprintings must be stored in a decimal register representing aggregate tube length in integral feet.

(*c*) Photo-electric limit switches must change the system over from electro-magnetic imprinting to wheel pulse counting, to measure the tail end.

(*d*) An additional count of 1 ft has automatically to be added to the footage register each time a tube is measured: this is to account for the unmeasured first foot at the commencement of the tube (see (*a*) above).

A block schematic showing how this was effected in a laboratory version is shown in Fig. 7. The three photo-electric limit switches PC1, PC2 and PC3 are shown in Fig. 6 and they control in the following manner. A tube passing along the conveyor covers first PC1 and then PC2. This combination causes the wheel to be lowered on to the tube. As the tube progresses along the conveyor PC3 is also covered, causing the write head to magnetize a spot on the tube 2 ft from its leading edge. When this magnetic spot appears under the read head normal imprinting ensues until PC1 is uncovered. This event coincides with the end of the tube being 2 ft from the read head. The circuit logic is so arranged that the next magnetic spot to appear beneath the read head switches in the tail end register. Pulses are accepted in this register until it is switched out of circuit again when PC2 is uncovered. Uncovering PC2 causes the measuring head to be lifted off (before the end of the tube actually reaches it: a safeguard against damage to the wheel).

These wheel pulses are stored in a modulo 16 register which has 'carry' facilities that feed into the footage register. The 'carry' pulse (if any) is stored and transferred to the footage register when PC3 is uncovered. Uncovering PC3 also initiates the +1 ft per tube requirements mentioned in (*d*) above.

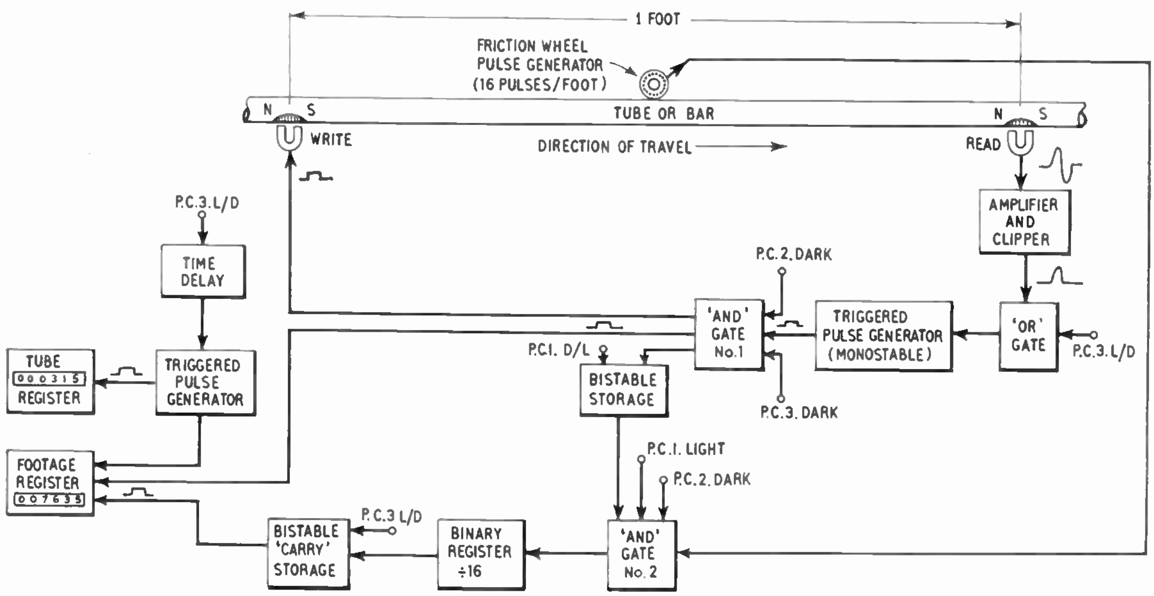


Fig. 7. Block schematic of the switching logic for the footprint-and-wheel system of measurement.

6. Expected Performance

Measurement errors can be classified in two groups: Those that occur once for every tube measured, and those that occur every foot (i.e. at each imprinting).

Those occurring in every tube can be caused by:

(*a*) Photocell operating tolerance.

(*b*) Measuring wheel switch on/off.

Those occurring every foot can be caused by:

(*c*) Variation in air-gap.

(*d*) Variation in tube speed.

(*e*) Electrical response times (can be neglected).

These are now considered in more detail.

(*a*) Previous experimental work has shown that photocells have an operating tolerance. In this application it is $\pm \frac{1}{8}$ in of tube travel. Three such cells are used to control the system and their combined error is not greater than 0·1% on mean tube length.

(*b*) The measuring wheel measures in units of $\frac{1}{16}$ ft. Depending on the position of the wheel when it is switched on and off, an error on the mean tube length of 0·2% could occur. On a large production run it is of a random statistical nature and will be self-compensating.

It is therefore probable that inaccuracy caused by (*a*) and (*b*) will not exceed ±0·1% for large random length batches.

(*c*) and (*d*). As can be seen from the oscilloscope traces, variation of tube speed and air-gap combine to vary the amplitude of signal (Fig. 5). Variations in amplitude causes errors in measurement because successive triggering of the print pulses is controlled from an arbitrary voltage level established in the reading head. The instant at which this occurs for any given speed/air-gap varies.

The amount by which the amplitude of the signal may be allowed to vary depends upon the degree of final accuracy required. To achieve an overall accuracy of ±1% a design target of ±0·8% has been used. Subtracting the ±0·1% utilized by (*a*) and (*b*) above, leaves ±0·7% acceptable variation for errors classified under (*c*) and (*d*).

It has been calculated that ±0·7% allowable variation, permits a 3 : 1 variation in signal amplitude (see Fig. 8). Figure 8(*a*) shows a low amplitude signal due to low speed and large air-gap, the tube has to travel $\frac{3}{12}$ in before reaching initiation voltage. Figure 8(*c*) shows a high amplitude signal due to high speed and small air-gap; the tube has to travel $\frac{1}{12}$ in to reach initiation voltage. If the amplitude ratio is 3 : 1 then the error about the average will be $\pm \frac{1}{12}$ in per foot or ±0·7%.

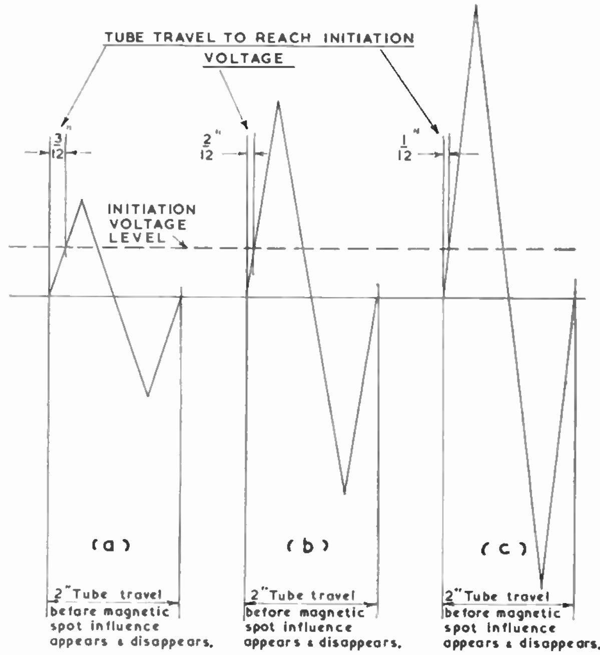


Fig. 8. Variation in tube travel to reach initiation voltage due to changes in signal amplitude. If the correct measurement is obtained when reading average signals, the error will be $\pm \frac{1}{12}$ in on either side giving an error of $\pm \frac{1}{12}$ in per foot or 0·7%.

Relating this to Fig. 5 the signal with a 0·060 in air-gap at 100 ft/min is 9 mV and the maximum signal with a 0·005 in gap at 300 ft/min is 28 mV giving an amplitude change of approximately 3 : 1 therefore the tube must be travelling at at least 100 ft/min with an air-gap of less than $\frac{1}{16}$ in if the error due to speed and air-gap is not to exceed 0·7%. It is expected that these conditions will prevail in practice.

7. Mechanical Layout for the Machine

A suggested layout for the machine is shown in Fig. 9. A box-like carrier arm is hinged via a bracket to the conveyor structure. Mounted centrally in the arm is a wheel which when brought into contact with the tube holds the read and write heads which are disposed on either side of it, $\frac{1}{32}$ in away from the surface of the tube.

There is a possibility that the tubes are slightly bent. However, since the measuring arm rides on the tube it is not expected that the air-gap will vary by more than $\pm \frac{1}{32}$ in over any one foot increment.

The pivot point about which the assembly hinges is chosen so that read and write heads are at right angles to the mean diameter of the tube when lowered on to it. Read and write head centres are fixed nominally at 12 in apart, the write head being adjustable so that optimum centres for greatest accuracy can be obtained by trial.

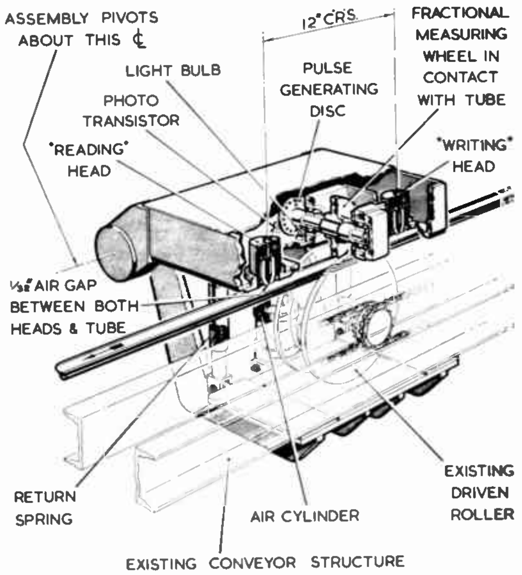


Fig. 9. Proposed apparatus for length measurement by electro-magnetic imprinting.

Raising and lowering of the carrier arm is achieved with an air cylinder. On the lowering stroke the air cylinder operates against two damping springs to ensure that the wheel comes into contact with the tube gently. The contact wheel has a circumference of 1 ft and is used for fractional measurement.

The electronic circuitry, footage and tube registers are housed separately, except for the photo-electric pulse generator and associated amplifier which is incorporated in the head assembly.

It is estimated that the total cost of the equipment would be in the order of £550.

8. Acknowledgments

The authors wish to thank Messrs. Tube Investments Limited for permission to publish this paper and acknowledge that the work was carried out under the Director of Development Engineering, Walsall. They would also like to pay tribute to and thank their colleague, George Powell, for his considerable contribution to the project, particularly in connection with the development work and experiments on the laboratory model.

9. Appendix: Potential Error Caused by End Loss

If one tube is measured, the error could be nearly 1 ft in 8 ft ($12\frac{1}{2}\%$).

If there are n tubes in a batch, then the error is $\frac{1}{2}n$ ft as $n \to \infty$. In this case we could eliminate end loss error by adding $\frac{1}{2}$ ft per tube measured to the length as registered.

We require to know how large n must be so that $\frac{1}{2}n$ added to the measured aggregate footage gives an answer that is better than $\pm 1\%$ of the true measurement.

The statistical distribution of the 1 ft 'tail end' range is rectangular: so when n is large, normal distribution is acceptable.

In this event it can be shown that

$$\sigma = \frac{1}{\sqrt{12n}} \qquad \ldots(1)$$

where σ is the standard deviation and a sample of r tubes is taken.

For normal distribution 99·73% of the population lies within $\pm 3\sigma$. To keep the range of 99·73% of batch samples within a range $\pm R$,

$$R = \pm 3\sigma \qquad \ldots(2)$$

Combining (1)+(2) we get

$$R = \pm \frac{3}{\sqrt{12n}}$$

If we now tabulate $R = \pm 3/\sqrt{12n}$, expressing R as a percentage of an 8 ft tube over the range of interest (i.e. where n is such that R approaches $\pm 1\%$) we have the results shown in Table 1.

Table 1

n	$R = \pm \frac{3}{\sqrt{12n}}$	% Error over 8 ft $= \pm \frac{R \times 100}{8}$
20	0·193	2·42
60	0·112	1·40
100	0·087	1·09
140	0·073	0·91

The batch size (n) would have to be 117 tubes or more to ensure that only 1 in 370 (i.e. 99·73%) batches of tubes were estimated to be outside $\pm 1\%$. This is of course, simply the random end loss error which has to be added to any other errors in the electromagnetic imprinting system. It becomes obvious then, that a measuring system based on 1 ft pitch imprinting plus calculated corrections based on adding 6 in per tube measured, would be too inaccurate to meet requirements where the batch size of 8 ft tubes might be as little as 20 tubes.

Manuscript received by the Institution on 23rd January 1963. (Paper No. 837.)

U.D.C. 621.391.83.4 : 621.397

On the Measurement and Interpretation of Non-linearity in a Television System

By

J. B. POTTER, F.R.M.T.C., B.Sc.†

Summary: A mathematical description of a general non-linear system is used to illustrate the scope of presently defined non-linearity measurements and to indicate how these may be extended. Present methods of interpretation of the results obtained from non-linearity measurements are discussed, leading to a more subjective assessment of the results being suggested as the basis of tolerances. Areas for further study in this field are also indicated.

1. Introduction

When a television system is tested by means of the various transient test signals generally used,[1] it is assumed *a priori* that a linear (equivalent) network is involved; that is, one which consists of linear, non-saturable, components only.‡ Where it is known that the system contains non-linear devices, as for instance in the case of a television transmitter using a vestigial-sideband system of modulation, this is taken into account by testing and specifying the response over a small enough region of the dynamic range of the device for its behaviour to approximate linear conditions.[2] When, however, the degree of non-linearity of the system as such is to be determined, the peak-to-peak amplitude of the test signals used (e.g., either the sawtooth[1] or staircase[3] waveforms) is such as to cover the full amplitude range of the system.

At the present time, it is only the non-linearity of the gain/amplitude characteristic of the system which is measured, this information being obtained from the system response to either the sawtooth or staircase waveforms mentioned above. The tolerance limits imposed on this non-linearity are chosen primarily with regard to maintaining the validity of the interpretation of transient response measurements made on the system. In general, no consideration is given to the effect which frequency/amplitude non-linearity has on the transient response measurements, yet in some situations this will be greater than that of gain/amplitude non-linearity. Further, no account is taken of the degree of subjective degradation of picture quality resulting from these non-linearities.

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

‡ Where the term linear is used in connexion with a system characteristic in the following, it should be interpreted as meaning that, in regard to the characteristic concerned, the system under test behaves as a linear (equivalent) network. The term non-linear is likewise used to denote that this requirement is not met.

It is the purpose of this paper to discuss the implications of extending the measurement of non-linearity in a video system, and the interpretation of the results obtained in objective and subjective terms.

2. A General Non-Linear System

Any system may be represented by a general equation of the form

$$E_{\text{out}}(a_i, s) = A_0 . A(a_i, s) E_{\text{in}}(a_i, s) \qquad \ldots\ldots(1)$$

at any instant in time t

where E_{in} and E_{out} are the input and output functions respectively

$A_0 . A$ is the transfer characteristic of the system

s is the complex frequency $= \sigma + j\omega$

and a_i is the amplitude level of the input.

Since we are concerned only with the base-band equivalent of a video system, we may, for convenience set $A_0 = 1$ as it represents a linear gain component of the system transfer characteristic only, and is irrelevant for these discussions. Of greater importance is the form of the function $A(a_i, s)$, and we will consider this in detail.

While the imaginary part of s (ω) is unrestricted as to the values it may have, a_i must remain within definite limits. For the video system the dynamic range must extend from a level representing black to a level representing white, and since only this range is of interest the convention may be adopted that black is represented by 0 and white by 1. Then $0 \leqslant a_i \leqslant 1$ defines the range of values available to a_i

For any normal video system it may be assumed that $\partial A/\partial s \neq 0$, i.e. the system transfer characteristic is frequency dependent. This follows from the fact that, in general, a video system will have the characteristics of a low-pass filter, if we look at its base-band equivalent. However, it will generally be possible to define restricted ranges of s within which $\partial A/\partial \omega = 0$ and the phase delay is constant. In a system having a linear frequency/amplitude characteristic, the form of $\partial A/\partial s$ is independent of a_i. It will, however, indicate both amplitude and phase variations in the transfer characteristic with frequency.

Consideration of $\partial A/\partial a_i$ leads to the following possibilities:

(a) $$\frac{\partial A}{\partial a_i} = 0$$

If this condition is satisfied then the system has a linear gain/amplitude response characteristic.

(b) $$\frac{\partial A}{\partial a_i} = h(a_i)g(s) \neq 0$$

where h and g are general functions. Under this condition, the system has a non-linear gain/amplitude response characteristic. This may be termed *static-non-linearity* as $h\ (a_i)$ is independent of s.

(c) $$\frac{\partial A}{\partial a_i} = h(a_i, s) \neq 0$$

i.e. $\partial A/\partial a_i$ is dependent upon (s), as the two variables are no longer independent. Such a system may be termed *dynamically non-linear* since the frequency/amplitude response characteristic is non-linear.

Hence we similarly find that for this case

$$\frac{\partial A}{\partial s} = g(a_i, s)$$

$$\frac{\partial A}{\partial \omega} \neq 0$$

i.e. the form of $\partial A/\partial s$ is dependent upon (a_i)

As previously discussed, it is normally assumed for convenience that condition (a) above is met in a television system. However, condition (b) is more likely because of the inherent static non-linearity of the active components used (valves, transistors, etc.). Dynamic non-linearity, defined by condition (c), may also occur; systems using frequency modulation or vestigial-sideband amplitude modulation are two examples where this type of non-linearity may occur.[4] It should be noted that both forms of non-linearity may occur in the same system but it is convenient to consider each separately.

2.1. *Static Non-linearity*

For convenience in considering this form of non-linearity it is assumed that $\partial A/\partial \omega = 0$ and that the phase delay is constant. From experience with practical video systems, it would appear that this assumption is justified provided we limit the range of $\omega/2\pi \leqslant 0{\cdot}2 f_B$ where f_B is the system bandwidth (i.e. $\omega/2\pi \leqslant 1$ Mc/s for 5 Mc/s system low-pass bandwidth). Hence, any test method designed to obtain information about static non-linearity should also be subject to like constraints.

From equation (1) we find that

$$\mathrm{d}E_{\mathrm{out}} = \left(A\frac{\partial E_{\mathrm{in}}}{\partial a_i} + E_{\mathrm{in}}\frac{\partial A}{\partial a_i}\right)\mathrm{d}a_i + \left(A\frac{\partial E_{\mathrm{in}}}{\partial s} + E_{\mathrm{in}}\frac{\partial A}{\partial s}\right)\mathrm{d}s \qquad \ldots\ldots(2)$$

which under the above conditions reduces to

$$\mathrm{d}E_{\mathrm{out}} = \left(A + a_i\frac{\partial A}{\partial a_i}\right)\mathrm{d}a_i, \qquad E_{\mathrm{in}} = a_i \quad \ldots\ldots(3)$$

Since a_i is the analogue representation of the grey-scale value of a particular area in the picture being transmitted, this equation represents the variation between the grey scale of the initial scene, and that of the reproduced scene. In other words, static non-linearity affects the grey-scale reproduction fidelity, i.e. the brightness-transfer characteristic of the system.

For convenience of notation, define

$$A(a_i, s) = \mathrm{f}(a_i) \quad \text{when } \frac{\partial A}{\partial \omega} = 0 \qquad \ldots\ldots(4)$$

and hence

$$E_o(a_i) = \mathrm{f}(a_i)E_{\mathrm{in}}(a_i) \quad \text{and} \quad E_{\mathrm{in}}(a_i) = a_i \quad \ldots\ldots(5)$$

Subjectively, the brightness sensation experienced by the observer is a logarithmic function of the stimulus.[5] It is therefore natural to apply the same logarithmic law to the technical video system and define a *system gradient* g_c such that for an ideal system

$$\frac{\log E_{\mathrm{out}}(a_i)}{\log E_{\mathrm{in}}(a_i)} = g_c = \text{constant} \qquad \ldots\ldots(6)$$

When $g_c = 1$, the system may be termed distortionless as the brightness-transfer characteristic will be such that despite any changes in absolute brightness, the contrast ratio (maximum to minimum brightness) remains unaltered between original and reproduced scenes.

If g_c has a value which is different from 1, uniform brightness distortion occurs since the grey-scale distinctions in the reproduced scene are uniformly compressed or expanded with respect to corresponding distinctions in the original scene. This is the same as saying that the contrast ratio of the reproduced scene is compressed or expanded with respect to that of the original scene. For $g_c > 1$ the objective effect in the reproduced scene is that of white expansion and black compression while for $g_c < 1$ the opposite occurs.

Combining equations (5) and (6) we find that

$$\mathrm{f}(a_i) = a_i^{(g_c - 1)} \qquad \ldots\ldots(7)$$

The shapes of f(a_i) for two values of g_c (= 0·9, 1·1) are given in Fig. 1. Note the large variation in f(a_i) which occurs in the black region in each case.

Hence provided f(a_i) can be measured and g is a constant, the actual value of g_c can be obtained. Practical measurement techniques and their interpretation will be discussed in later sections of this paper.

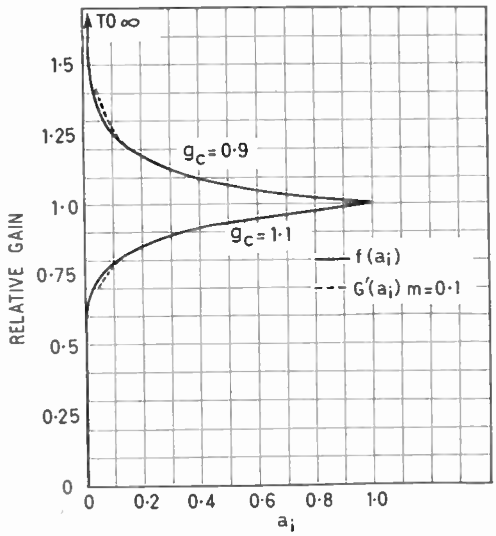


Fig. 1. f(a_i) and $G'(a_i)$ for a system of gradient g_c = 0·9 and 1·1.

Video systems consist of a large number of non-linear elements as previously discussed, so that we must accept the fact that the system gradient will vary from point to point throughout the range of a_i values allowed. Under these conditions we may consider the system gradient to consist of two components, namely

$$g = g_c + g_r \qquad \ldots\ldots(8)$$

where g_r is a non-systematic variation component of g so that f(a_i) varies in some non-systematic fashion about the values given by equation (7).

2.2. *Dynamic Non-Linearity*

Outside the region where $\partial A/\partial\omega = 0$, it is assumed for generality that the system exhibits dynamic non-linearity; that is both $\partial A/\partial s$ and $\partial A/\partial a_i$ are dependent in form on both a_i and s. Looking at equation (2) again

$$\mathrm{d}E_{\mathrm{out}} = \left(A\frac{\partial E_{\mathrm{in}}}{\partial a_i} + E_{\mathrm{in}}\frac{\partial A}{\partial a_i}\right)\mathrm{d}a_i + \left(A\frac{\partial E_{\mathrm{in}}}{\partial s} + E_{\mathrm{in}}\frac{\partial A}{\partial s}\right)\mathrm{d}s$$

it can be seen that there are a number of ways of testing the system for this type of distortion. These will include quantized amplitude steps with frequency sweeping or vice-versa. It is, however, apparent that no matter how the measurements are made it is necessary to define the conditions under which the comparison response of the system is obtained. If quantized frequency steps are used then the response with varying a_i obtained as the static non-linearity response can be the comparison response, and the variation from this frequency taken as a measure of dynamic non-linearity. If quantized amplitude steps are used, however, a new comparison response must be defined, possibly the response for $a_i = \frac{1}{2}$.

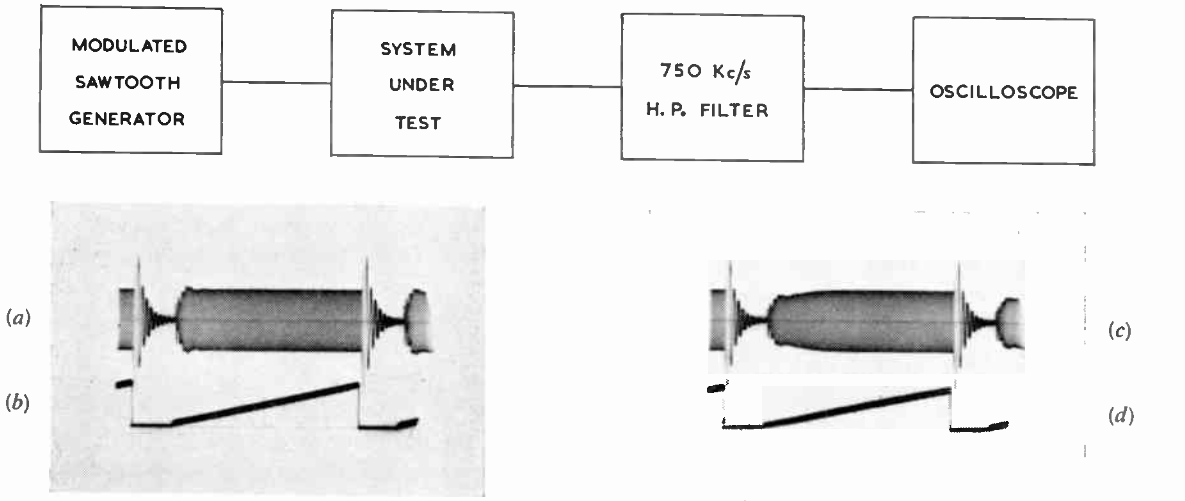
Subjectively it is found that dynamic non-linearity manifests itself as a change in the sharpness of transitions in the reproduced scene, depending on the position of the transition within the grey scale. No attempt has been made as yet to obtain quantitative results to determine the maximum allowable dynamic non-linearity. However, qualitatively it would appear that maximum sharpness in the mid-grey to white region of the grey scale is desirable if any dynamic non-linearity at all can be tolerated. Certainly the black areas are the least critical as to sharpness, while the subjective reaction to 'bloomed' white areas which often occur due to camera defects etc., are normally very noticeable.

However, all such comments must remain conjecture until further experiments, designed to statistically determine subjective limits of acceptability, can be performed.

3. Methods of Measurement

At present, only static non-linearity is normally measured in video systems and this information is commonly obtained by either of two techniques.[1,3,6,7] These are based on the use of a line sawtooth waveform modulated with a low amplitude sine wave, (C.C.I.R. test signal 3) and a 5 to 10 step line staircase waveform, the number of steps selected being dependent on noise conditions. In both cases the frequency range of the waveform is restricted to $0{\cdot}2 f_B$ so as to meet the $\partial A/\partial\omega = 0$, constant phase delay requirement discussed above.

Briefly, the sawtooth waveform has superimposed upon it a low-amplitude sine wave of frequency $0{\cdot}2 f_B$. This waveform is passed through the system after which a simple high-pass filter is used to separate out the sine wave. The method and waveforms involved are illustrated in Fig. 2. In the case of the staircase, the waveform, after passing through the system, is differentiated and shaped to an approximate sine-squared form by means of a low-pass shaping network[7] with a cut-off frequency below $0{\cdot}2 f_B$. For the ten-step staircase shown in Fig. 3, a network time-constant of 1·25 µs was used (Australian 625-line system).



(a) Recovered 1 Mc/s modulation.

(b) Sawtooth before detector.

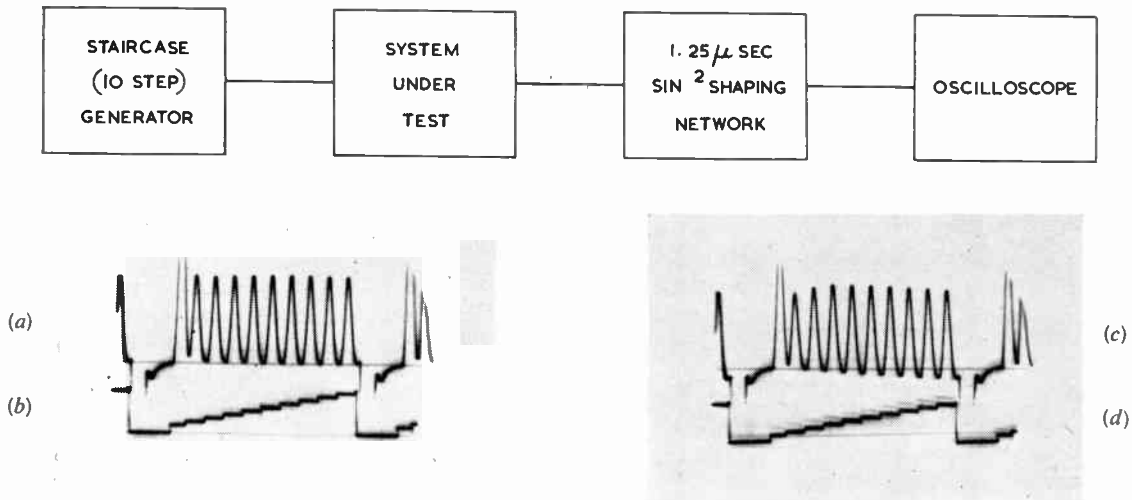
(c) Recovered 1 Mc/s modulation after passing through a non-linear device.

(d) Sawtooth before detector.

Fig. 2. Block schematic and waveforms involved in the measurement of $G'(a_i)$ using a modulated sawtooth.

In both cases the final result is an approximation to $f(a_i)$ which we may denote by $G'(a_i)$. The slight difference between these two functions comes about because of the finite amplitude of the sine wave modulation and the step amplitude respectively. In other words, in the limit $G'(a_i) \rightarrow f(a_i)$ for infinitely small sine wave amplitude and step amplitude. These relationships are developed and discussed in Appendix 1.

A further method, capable of accurately determining the value of system gradient (g_c) is described in Ref. 8. Two exponential voltage generators, one of fixed exponent, the other variable, are used to produce line frequency waveforms. The two waveforms are fed via the system under test to the Y and direct to the X plates of a c.r.o. respectively. The variable exponent is adjusted to obtain a straight line on the screen of the oscilloscope, this condition indicating that the wave-



(a) Shaped step transitions.

(b) Original step waveform.

(c) Shaped step transitions after passing through a non-linear device.

(d) Step waveform before shaping.

Fig. 3. Block schematic and waveforms involved in the measurement of $G'(a_i)$ using a staircase.

forms to X and Y plates are identical. As shown in Appendix 2, the value of g_c for the system can be obtained from these, provided the two exponents are known. Such a method could give important supplementary information about a system, particularly if constant gradient correction networks are incorporated.

Measurement of dynamic non-linearity can be performed in a number of ways, but since the modulated sawtooth and staircase waveforms are already in use for static non-linearity measurement, we will confine our discussion to these. Many equipments which produce a modulated sawtooth waveform also provide for the switching, in steps, of the modulation frequency from 1 Mc/s to 6 Mc/s. Such a waveform is very suitable for the measurement of frequency response distortion with amplitude but the phase response variations can only be obtained with some difficulty. Hence only part of the dynamic non-linearity of the system can be measured in this way. Within this limitation, however, Mueller[4] has used the modulated sawtooth technique to investigate both static and dynamic non-linearity measurements, the latter measurements being performed on frequency-modulated radio links. The results obtained illustrate quite forcibly the importance of dynamic non-linearity measurements on a system.

If we carry to its natural conclusion the idea that. in a television system, transient response tests are more indicative of what happens in practice than the steady-state type of test, then the staircase waveform should provide more meaningful dynamic non-linearity information than the modulated sawtooth.

Extension of the staircase test method previously described could be made for this purpose by shaping the rise of each step to have a sine-squared form of $2T$ characteristic period. A pulse-and-bar display could then be produced by displaying the step and its derivative superimposed on the oscilloscope screen. In this way a K-rating[7] for each step could be determined. The variation of K-rating with step position would then completely describe the dynamic non-linearity of the system.

4. Static Non-Linearity Characterization

The present method of interpreting static non-linearity is detailed in a C.C.I.R. Recommendation.[6] Briefly, this interpretation necessitates the determination of the maximum and minimum values of $G'(a_i)$ only, the ratio being required to lie somewhere between 0·8 and 1·0; if the system meets these limits it is deemed acceptable. But such an interpretation fails to consider the importance of the shape of $G'(a_i)$.

As discussed above, the shape of $G'(a_i)$ is of importance from the subjective point of view, a fact which has been illustrated by Mueller.[4] For example, he reports that in subjective tests, using a linear black to white variation in $G'(a_i)$ (which has a gradient variation component plus a non-systematic component present), minimum to maximum $G'(a_i)$ ratios from 0·7 to 0·8 are acceptable. However, when the non-systematic variation alone was present (here a white crushing effect) values of 0·8 to 0·9 defined the acceptable limit.

Because the shape of $G'(a_i)$ can be due to both system gradient and non-systematic distortion through the system, it would be convenient to specify limits for each of these separately. If this is done then some means of separating the two components must be found. This can be done graphically using $G'(a_i)$. Another possible method is to use a standard gradient correction network to remove the gradient distortion, and then measure the g_c value from this correction network, using exponential waveforms as discussed. The non-systematic variation remaining could then be measured in the same manner as the whole static non-linearity is determined at present. This may not of course be the final answer, but it would be a good starting point for further investigation.

Having defined two components, it also becomes necessary to set limits which are acceptable for these components. Again, only a starting point can be indicated although some evidence from experience has influenced the selection of the following suggested limits:

(a) System gradient g_c is to lie between 0·9 and 1·1. Such limits are in keeping with the results of Mueller although they appear conservative when compared with the variations in g_c which occur due to receiver adjustment by viewers.

(b) Non-systematic distortion measured as at present, to be limited to a ratio of $G'(a_i)_{min}$ to $G'(a_i)_{max}$ between 0·9 and 1·0. This limit is mainly suggested by the visibility of these effects in the white region.

5. Conclusions

In the above discussion it has been shown that present methods of specifying the non-linear performance of video systems are inadequate for two reasons. These are:

(a) Only the static non-linearity (i.e. gain/amplitude non-linearity), of the system is measured, dynamic non-linearity (i.e. frequency/amplitude non-linearity) being ignored; and

(b) the limits specified for static non-linearity do not take into account the subjective end-result of these limits.

Some suggestions have been made to overcome these objections. Dynamic non-linearity can probably be best measured by use of a step-waveform, shaped so as to allow its use as a 'pulse and bar' waveform at each amplitude change. The degree of non-linearity could then be characterized by the variation of K-rating of the 'pulse-and-bar' waveform with position in the range of the grey scale. The use of the modulated sawtooth waveform with variable modulation frequency is another possibility, although less attractive than the first. No attempt has been made to define limits for this form of non-linearity apart from commenting that maximum sharpness in the mid-grey to white region is probably a requirement.

Separation of the static non-linearity into a system gradient component and a non-systematic component has been put forward as a means of characterizing this form of non-linearity in a way which has subjective significance. A system of limits based on this can then reflect the less noticeable nature of the system gradient component, while allowing tighter control of the non-systematic component to be exercised. Tentative values for these limits are suggested as a starting point for further investigations.

6. Acknowledgments

The author wishes to thank the Engineer-in-Chief of the Australian Post Office for permission to publish this paper, and Mr. A. J. Seyler for many helpful discussions.

7. References

1. A. J. Seyler and J. B. Potter, "Waveform testing of television transmission facilities", *Proc. Instn Radio Engrs Aust.*, **21**, p. 470, July 1960.
2. S. F. Brownless, "Delay and transient problems in television broadcasting", *Proc. Instn Radio Engrs Aust.*, **21**, p. 253, April 1960.
3. L. E. Weaver and I. J. Shelley, "Measurement techniques for television broadcasting", *J. Television Soc.* **9**, No. 12, p. 468, October–December 1961.
4. J. Mueller, "Uber die nichtlinearen Verzerrungen von Fernsehleitungen", *Archiv Elektr. Ubertrag.*, **11**, p. 485, December 1957.
5. D. G. Fink, "Television Engineering", Chap. 18, p. 44, 2nd Edition, (McGraw-Hill, New York 1952).
6. C.C.I.R. Recommendation No. 267 (Los Angeles 1959).
7. N. W. Lewis, "Tentative Requirements for the Transmission of 625-Line Television Signals", G.P.O. Research Report No. 20661, London, February 1961.
8. "Method of Measuring the Numerical Characteristics of Gamma Correctors", C.C.I.R. Document 124-E, March 1959.

8. Appendix 1

Interpretation of the Modulation Envelope of a Modulated Sawtooth and the Differentiated Pulse Heights of a Staircase in a Video System

Consider a system characterized by equation (5)

$$E_{\text{out}}(a_i) = \text{f}(a_i)a_i \qquad \ldots\ldots(9)$$

where a_i, the input level is restricted in the range

$$0 \leqslant a_i \leqslant 1, \quad (0 = \text{black and } 1 = \text{white})$$

and

$$E_{\text{out}}(a_i) = 0, \quad a_i = 0$$
$$E_{\text{out}}(a_i) = 1, \quad a_i = 1$$

Further, we may define an "ideal" video system such that

$$\frac{\log E_{\text{out}}(a_i)}{\log(a_i)} = \text{constant} = g_c \qquad \ldots\ldots(10)$$

Combining (9) and (10) we find that

$$\frac{\log \text{f}(a_i)(a_i)}{\log(a_i)} = g_c$$

which can be solved to give

$$\text{f}(a_i) = a_i^{(g_c - 1)} \qquad \ldots\ldots(11)$$

Such a function also meets the boundary conditions imposed on $E_{\text{out}}(a_i)$ and a_i and is therefore an acceptable solution for $\text{f}(a_i)$.

We may now apply to the system a sawtooth which goes from black to white (and defines the value of a_i) with a superimposed h.f. sine-wave E_m of amplitude $m \ll 1$. Then the amplitude of the sine wave at the system output will be

$$E_{m(\text{out})} \simeq g_c . m . \text{f}(a_i) \qquad \ldots\ldots(12)$$

which is used in the normalized form obtained by dividing (12) by g_c. This will become exact in the limit as $m \to 0$. However, in practice m may not be less than 10% of the peak sawtooth amplitude.

Hence the observed waveform is

$$E_{m(\text{out})} = m . G'(a_i) \qquad \ldots\ldots(13)$$

where $G'(a_i)$ is the normalized form of $G(a_i)$, the apparent gain/amplitude transfer function of the system. Normalization is necessary since although $\text{f}(a_i) = 1$, $a_i = 1$ by definition, the same is not true of $G(a_i)$ but in making measurements this cannot be taken into account.

When $\text{f}(a_i)$ is of the form in equation (11)

$$G'(a_i) = \frac{\left(a_i + \frac{m}{2}\right)^{g_c} - \left(a_i - \frac{m}{2}\right)^{g_c}}{\left(1 + \frac{m}{2}\right)^{g_c} - \left(1 - \frac{m}{2}\right)^{g_c}} = \frac{G(a_i)}{G(1)} \qquad \ldots\ldots(14)$$

Curves for $\text{f}(a_i)$ and $G'(a_i)$ for $g_c = 0{\cdot}9$ and $1{\cdot}1$, $m = 0{\cdot}1$ are shown in Fig. 1. From this it can be seen that $G'(a_i)$ closely approximates $\text{f}(a_i)$ for a_i values between 0·1 and 1·0.

$G'(a_i)$ cannot be evaluated below $a_i = m/2$, but since the modulation envelope is generally disturbed in this region, this is not of any great importance.

The above analysis may also be applied to the staircase waveform by assuming that it consists of a series of steps of height m, each step being symmetrically placed about a series of values of a_i such that

$$a_i = \frac{m}{2}, \frac{3m}{2}, \ldots, 1 - \frac{m}{2}$$

Hence a ten-step waveform will provide ten samples of the function $G'(a_i)$ between black and white, with the first value at $a_i = 0{\cdot}05$, the last at $a_i = 0{\cdot}95$ and with $m = 0{\cdot}1$.

It will be noticed that in this latter case, white is no longer usable as a reference point for determing the effective g_c value of the system. However, $a_i = 0{\cdot}95$ can be used to represent the white reference point with an error of the order of 1% for g_c close to 1.

Appendix 2

System Gradient Measurement by Means of an Exponential Waveform

Assume the test voltage is of the form

$$E_{\text{in}} = A\,\mathrm{e}^{-\beta t} \qquad \ldots\ldots(15)$$

and the transfer function is of the form

$$E_{\text{out}} = \alpha(E_{\text{in}})^{g_c} \qquad \ldots\ldots(16)$$

then

$$E_{\text{out}} = \alpha A^{g_c}\,\mathrm{e}^{-\beta g_c t} \qquad \ldots\ldots(17)$$

Let us compare the voltage of (17) with

$$x = x_0\,\mathrm{e}^{-\lambda t} \qquad \ldots\ldots(18)$$

so that E_{out} is applied to the Y plates of a c.r.o. and x is applied to the X plates.

A straight line will be formed at 45 deg across the screen when

$$x = E_{\text{out}}$$

i.e., when $x_0\,\mathrm{e}^{-\lambda t} = \alpha A^{g_c}\,\mathrm{e}^{-\beta g_c t}$

Now $x_0 = \alpha\, A^{g_c}$ can be adjusted by a straight gain variation in the x path.

Hence a straight line occurs when

$$-\lambda t = -\beta g_c t$$

or

$$g_c = \frac{\lambda}{\beta} \qquad \ldots\ldots(19)$$

Hence provided λ and β are known, g_c for the system can be readily obtained. Since exponential waveforms are required, simple RC-circuits may be used to generate the waveforms and hence β and λ are given by the time constants of these circuits. By using identical C values, the ratio of R values gives the information required. Measurement error and methods are discussed in Ref. 8.

Manuscript received by the Institution on 25th October 1962. (Paper No. 838/T23.)

"SIGNAL PROCESSING IN RADAR AND SONAR DIRECTIONAL SYSTEMS"

(With special reference to systems common to Radar, Sonar, Radio Astronomy and Ultrasonics.)

A symposium on the above subject, sponsored jointly by the Institution and the Electrical Engineering Department of the University of Birmingham, will be held at the University of Birmingham from 7th to 10th July 1964. Its purpose is to bring together those working in the fields of Radar, Sonar, Radio Astronomy and Ultrasonics to enable them to discuss common aspects of the signal processing techniques as applied in these different fields.

The development of signal processing techniques in these various applications has continued independently despite the common work involved, and it is clear that a useful purpose would be served in bringing such experts together to exchange new ideas and techniques. The scope of the symposium will include the following topics in so far as they are concerned with directional systems:

- Non-linear signal processing and correlation systems
- Directional pattern synthesis including aperture synthesis
- Electronic scanning systems
- 'Within-pulse' scanning
- Multiple beam-forming techniques
- Digital processing applied to directional systems
- Directional systems for automatic tracking
- Wide-bandwidth directional systems
- The effect of phase-coherence in directional systems
- Transient effects in directional systems
- Pulse compression

Contributions are invited on any of the above, or related topics, for consideration for inclusion in the programme of the symposium and offers of papers from overseas will be especially welcomed. Summaries should be submitted to the Institution as soon as possible. It is intended to preprint all papers; an announcement about subsequent publication will be made later. Details relating to registration and accommodation in a University Hall of Residence will be circulated shortly.

Electronics at the Royal Society Conversazione

Electronics research now features prominently at the annual Conversazione of the Royal Society, and this year three items were particularly noted at the press preview.

Topside sounding of the ionosphere by means of the Canadian satellite *Alouette*† was illustrated in an exhibit by Dr. J. W. King of the Radio Research Station (D.S.I.R.). The satellite telemeters information derived from its topside sounder to data acquisition stations all over the world. The telemetered data are recorded on magnetic tape and sent to the Radio Research Station at Slough where they are converted to 'ionograms' for analysis. This has revealed much that was previously unknown about the topside of the ionosphere. The van Allen belts, which are regions of trapped energetic particles which travel back and forth along certain lines of the earth's magnetic field, apparently interact with the atmosphere to produce a wide variety of phenomena. Another diagram showed the remarkable extent to which the charged particle density in the upper ionosphere is controlled by the earth's magnetic field. A third diagram showed how the so-called 'scale height' of the electron distribution, which depends on the temperature of the ions and electrons, varies with latitude. There appears to be a marked variation of this scale height with latitude which is not yet understood.

The Services Electronics Research Laboratory (Mr. R. F. Broom, Mr. C. H. Gooch, Dr. C. Hilsum and Dr. D. J. Oliver) showed a gallium arsenide optical maser which sends out an intense beam of infra-red radiation when electrical current passes through it. It works at the temperature of liquid air, 90° K, and at current densities above 2000 amps/cm². The gallium arsenide is a small cube of edge 0·4 mm with contacts attached to top and bottom sides. When current passes through the cube, radiation is emitted from a line about half way up it. About 50% of the current is converted into radiation of the wavelength 8500 Å, i.e. outside the visible spectrum. The line of light is, however, effectively brighter than the sun, for its brightness is greater than 1 MW/cm² whereas the sun's brightness is only 10 kW/cm². This device is the most efficient instrument we have for converting electricity into radiation. It can be pulsed at more than 1000 Mc/s, and therefore should be useful in optical communication systems.

The Direct Conversion Group at the Atomic Energy Research Establishment, Harwell (Mr. G. Rice and Dr. J. Myatt), gave details of some of the work on the possibilities of thermionic emission. It was Edison who in 1883 discovered that a solid, when heated to a high temperature (1000–2000° C), emitted electrons. Unfortunately, fundamental problems, basically concerned with cathode materials, seemed to indicate that the diode would never be capable of generating useful quantities of electrical power. However, towards the end of the last decade it was realized that these were not insurmountable and that the power output from diodes could be increased several hundred-fold. Today the problems that remain are mainly technological ones and, if these can be solved, there seems to be no reason why diodes with nuclear heated cathodes should not be able to convert nuclear power into electrical power with efficiencies of up to about 25%. Diodes could work in conjunction with the conventional steam generators to increase the efficiency of the nuclear power generating stations by between 8% and 15%. At Harwell, use is made both of electrically heated cathodes in the laboratory and also of nuclear heated cathodes using the reactor—PLUTO—at Harwell. In the case of the former a working example was demonstrated and diagrams of the latter type were shown.

† R. C. Langille and J. C. W. Scott, "The Canadian Defence Research Board topside sounder satellite", *J.Brit.I.R.E.*, 23, No. 1, pp. 61–8, January 1962.

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **July 1963**

1963 July	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	−130·4	−131·6	+19
2	−130·5	−130·0	+17
3	—	−130·9	+19
4	—	−129·8	+20
5	—	−129·5	+21
6	−130·0	−130·0	+21
7	—	—	−16
8	—	−129·8	−17
9	−129·8	−130·3	−17
10	−130·1	−130·4	−15
11	−129·9	−129·7	−18
12	−130·3	−130·3	−15
13	−130·2	−129·8	−16
14	−129·3	−129·7	−13
15	−130·2	−130·6	−12
16	−129·8	−129·5	−12
17	−129·2	−129·8	−11
18	−130·5	—	−11
19	−130·2	−130·8	−10
20	−130·7	−130·8	−12
21	−130·6	−130·7	−10
22	−131·1	−131·2	−10
23	−129·3	−130·3	−10
24	−130·4	−131·6	+17
25	−129·9	−130·2	+19
26	−128·8	−129·8	+20
27	−129·0	−129·4	+20
28	−129·5	−131·3	+18
29	−131·5	−132·1	+17
30	−130·2	−131·8	+19
31	−130·3	−130·0	+20

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

U.D.C. 374.64: 621.37/9

The Use of Projects in the Practical Training of Professional Engineers for the Telecommunications and Similar Industries

By

L. P. GRICE, B.Sc.,†

V. J. LEWIS, B.Sc.†

AND

A. D. MATTHEWS, B.Sc.‡

Presented at a meeting of the Education Group in London on 24th April 1963.

Summary: The paper is specifically concerned with training in the telecommunication engineering industry, although the methods described are of general application.

The particular problems which arise in the training of professional engineers as opposed to craftsmen are discussed, and it is shown how 'projects' can be devised that take account of the relatively short time available for practical training and require the student to exert himself and exercise initiative yet at the same time drawing his attention to important practical details.

Typical examples are given covering a wide range of different departments, and a special merit of the training methods indicated is that they can be operated under normal industrial conditions without any great requirement for additional staff or facilities.

1. Introduction

It is generally accepted that the electrical engineer specializing in the heavy current field must follow some system of planned practical training before accepting responsibility as a professional man. In the light current side of the profession, however, this has not been so widely accepted and opinion is divided between those who regard practical training as essential and those who think it a waste of time.

While accepting that there will always be some men, especially those going into some types of research work, who will not be in immediate need of practical training, it is the authors' opinion that most men entering the profession would be well advised to obtain a more varied experience of the different facets of industry than can be obtained by going into direct employment.

The traditional way of obtaining this experience is by undertaking either a graduate or student apprenticeship. The graduate apprenticeship normally lasts two years and usually follows immediately after a university degree course in engineering or possibly, physics.

Although some engineers still qualify by means of part-time, or even evening study, this route has become progressively more difficult and it is likely that the majority of student apprentices will take 'sandwich' courses in the future. The structure of the sandwich varies considerably, but the total length of the course is usually between four and five years. The total duration of the industrial periods is usually 2–3 years, so that the time spent in practical training is approximately the same in both the student and graduate apprenticeship.

Programmes of industrial training vary greatly, but a scheme in the light electrical industry occupying a total period of two years would usually have the following general form:[1]

Training	Location	Period
(*a*) Basic workshop practice	Training school or workshops	3–4 months
(*b*) General mechanical and electrical training	Assembly and processing, inspection, test and development	9–12 months
(*c*) Directed objective training		9–12 months

The last period of training would be planned having in mind the student's particular interest and the field in which it is anticipated that he will finally seek employment. A professional trainee in a manufacturing concern might spend this last period in one of the following fields:

Research and development
Design
Manufacture (production and maintenance)
Installation and commissioning
Commercial engineering

The first practical training schemes for graduates were little more than adaptations of the existing craft apprenticeships. This, however, is not a satisfactory solution because:

(a) The graduate needs a greater intellectual challenge.

† The General Electric Company (Telecommunications) Ltd., Coventry, Warwickshire.

‡ Lanchester College of Technology, Coventry, Warwickshire.

(b) He must be given an opportunity of exercising initiative.
(c) The time available for his training is too short for him to spend long periods watching.
(d) The objectives of his training are quite different.

Most graduate or student apprentices on entering industry are naturally anxious to make progress as fast as possible and they are intolerant of any process which seems to them time-wasting or unproductive in knowledge. The supervisor of a production department whose own background is frequently more akin to that of the trade apprentice than the professional, often feels quite correctly that the time available is too short for the trainee to acquire the practical ability to undertake any work of an advanced nature. Unless special arrangements are made, the student may thus only be given work of a trivial or repetitive nature and will be permitted only to watch the more complicated processes. The result is inevitably to breed in the trainee a sense of frustration and boredom. The danger of such boredom cannot be over emphasized as the enthusiasm which most trainees bring into industry is easily lost and lack of activity leads to apathy. A way of overcoming this problem is described in detail in a paper on techniques of graduate training by Marples, Radford and Reddaway.[2] However, very few organizations would be able or willing to provide all the facilities necessary for this approach and it is considered that the real value of the authors' solution is that it is capable of operating under normal industrial conditions without any significant requirement for additional staff or facilities. To utilize his more academic outlook to advantage, the trainee is given specific projects which involve finding out for himself the details of practical operations, and examining them in the light of his theoretical knowledge.

The value of project training is emphasized in the Brit.I.R.E. report on "The Practical Training of Professional Radio and Electronic Engineers".[3] The application of project training to groups of graduates for machine shop practice has been described by G. H. Russell,[4] but the methods described here can be applied on an individual basis and also more extensively.

2. Training Objectives

The following points should be borne in mind:

(a) The trainee does not normally need to become expert at the operations he will study, and in any event there is no time for this. What he does need is an understanding of the possibilities and limitations of the materials and techniques involved.

(b) In general, however, some practical experience of the work involved is needed so that these possibilities and limitations should be fully understood.

(c) The trainee should be set a task which makes him exert himself to the limit of his abilities and which lets him realize that the problems of industry are, in their own way, just as fascinating and complex as those of a more academic nature. He must be encouraged to use his own initiative.

(d) He should be required to mix with personnel at the shop floor level, and perhaps one of the most important points that he must learn is how to get on with the shop floor supervisor. Many graduates, for example, because of a lack of knowledge of the background and outlook of production staff, fail to obtain the fullest co-operation.

(e) Where possible, practical methods should be related to some theoretical concept which the trainee understands.

3. Scope of Training

The types of department that the trainee will visit during his training in industry can be grouped under six headings.

(a) Workshops and assembly department, including servicing and finishing department.

(b) Test and inspection departments.

(c) Administrative departments, e.g. production control.

(d) Design or drawing offices.

(e) Development or research departments.

(f) Production engineering and methods department.

The departments listed in (a), (b) and (c) can be considered together as a similar approach is required in devising projects for them. Groups (d), (e) and (f) need individual attention and projects for these are explained in Sections 4–7.

A training scheme for each of the different types of work involved will now be considered.

4. Training in Production, Test and Administrative Departments

These departments will normally prove most difficult as their training value is not immediately obvious to the trainee because of his academic rather than practical outlook. The approach to be described consists of setting the trainee a definite 'project' to undertake in each department, this being far more than just a series of questions to which he must find the answers. It is a scheme which calls for close co-operation between those responsible for organizing the training and the supervision of the department concerned. In many cases, however, the supervisors have been glad to have a definite scheme for students under their charge who had formerly been somewhat of an embarrassment.

Every department has its own set of circumstances, and projects must be produced separately in every case and kept under constant supervision and revised as circumstances may demand. A possible method of undertaking this is described later.

4.1. *Method of Project Design*

To devise a project the following general scheme may be used:

(a) The scheme is explained in detail to the supervisor of the department and his co-operation is obtained.

(b) The work of the department is carefully analysed and points which the supervisory staff consider to be of particular importance are noted.

(c) An overall assessment of the knowledge which the student is expected to gain is then made in consultation with the supervisor, bearing in mind special features of the work as well as the general picture.

(d) A project is then devised which, while testing the student's initiative, will require him to cover all the points in (c), and provisional timing is made for each part of the project.

(e) The scheme is explained to all members of the department with whom the trainee would be concerned. It is essential to the success of the scheme that the trainee should not be regarded as a "spy" and that the reason for his presence should be understood by everyone. The support of the operators and supervisory staff is required at all stages.

(f) In order to direct the trainee's attention to particular points, and to avoid undue waste of supervisors' time, a number of headings to be included on the report are given in the project sheet. These are merely designed as a guide to direct the trainee's attention to points which he might otherwise overlook.

(g) It will be necessary for the trainee to undertake some reading on the subject concerned, and references to a suitable bibliography are given. It is expected that the man will undertake such general reading or theoretical work in his own time.

(h) The greatest importance is attached to the report written on the project and this should always be read by both the supervisory staff of the department concerned and the training department.

(i) The progress of the first men attempting a new project is watched with great care. The success of the exercise is discussed with the student and the people involved at all levels, and modifications made to the projects and timing where necessary.

In the following examples, production, test and administrative departments will be dealt with in turn. These departments could all refer to one product so that the entire picture of the manufacture of that product could be seen, but it is appreciated that in any practical training scheme where large numbers are involved, only a few trainees may be able to follow this route. It is a feature of this system of projects that they are designed to give a general outlook on the problems of production, test and administration which is not dependent on the actual departments concerned. It is thus not necessary for each trainee to visit every production floor or every test room. One or two representative departments may be chosen, and the understanding of the problems involved may be applied elsewhere. Hence the calls on the trainee's limited time may be kept within reasonable bounds.

4.2. *Production Departments*

4.2.1. Machine shops

A knowledge of machine shop techniques and the difficulties and limitations of these departments is of great importance to the engineer. The trainee should have practical experience in machining techniques prior to undertaking a machine shop project. This is best given in a training school, but where this is not possible a month could be spent in the actual machine shop making test pieces. The actual project in the machine shop should involve the problems of quantities and machine utilization as well as give an understanding of what is easy to produce and what causes difficulty. An example of a project designed for a machine shop is given in Appendix 1.

4.2.2. Assembly departments

These may be subdivided into small quantity and mass production assembly sections. The approach will be different in every case, but first it must be decided to what extent the trainee will be able to gain practical experience. Repetitive operations must be avoided unless they have some definite instructional objective.

In the mass production assembly shop where every operation is organized in detail there may be little opportunity for giving the student any personal responsibility for production, or any extensive practical training, so he must be asked to report on specific points concerning the operation of the assembly unit.

These may, for example, be under the following headings:

Lay-out. To study the disposition of the machines and operators with reference to efficiency and the handling of materials.

Process. To investigate the processes involved in the various operations and their relationships to one another.

Technical. To study the technical background of the product and the production processes.

Yield and Economy. To investigate the reasons for 'rejects' and the methods by which improvements in economy can be made.

Process Control. To investigate the way in which quality is maintained and the steps taken to avoid a deterioration in standards.

At each stage an attempt should be made to draw the trainee's attention to some general aspect of industrial life, e.g. the problems of piecework payment, safety regulations, etc.

An example of a training project for a typical mass production unit is given in Appendix 2.

In the case of a small quantity production department some of the scheme for a mass production department will apply, but the intermittent nature of the work may mean that a detailed critical examination of the processes involved, may be impossible.

On the other hand the nature of the work is often such that the trainee, under the supervision of a junior foreman, may be personally responsible for the organization of the production of a small order. He will thus learn a very great deal about production problems in a short while.

4.3. *Test and Inspection Departments*

To reap the full advantage of experience in a production department the student should have a period working on test or inspection. With simple tests of a mechanical nature, excessive repetitive work should be avoided. With more complex electrical testing a student can test large numbers of similar equipments and still obtain useful experience. In each case the trainee should be given a project requiring the investigation of a specific problem. He should be expected to explain the technical aspects of the work involved and to undertake any reading that may be necessary for this purpose. The answers given to his technical queries from operators engaged in routine testing may well be unsatisfactory and he must seek information elsewhere. The training department must be able to find a source of information if the man is unable to do this for himself.

A special feature of test department projects is that the trainee can examine the work in the light of his previous technical knowledge and apply this in practice. Appendix 3 gives a typical example of this.

4.4. *Administrative Departments*

The system of projects may equally well be applied to 'paper work' departments. A large number of people in industry are engaged in various administrative sections, and to obtain a broad overall picture it is necessary to have some idea of the work that they do. This point is however, often not apparent to the trainee, who may feel that his interests lie purely in technical matters. It is hence most important that a project is drawn up to give him a definite aim for his stay in the department, and to avoid the superficial idea that there is "nothing to it but a lot of form filling".

An example for a scheme for a production control department is given in Appendix 4.

It is inevitable that as he progresses in his profession the trainee will find himself more and more involved in work of an administrative nature, and an early introduction into the principles of such work will be most valuable.

5. Training in Production Engineering and Methods Departments

Specific projects are difficult to prepare in advance for such departments. The solution is to obtain the co-operation of an engineer in the department who is willing to explain his own projects to the trainee and involve him in each stage of an investigation. A clear report on the work done and background reading would always be required from the trainee. Examples of typical projects are:

(a) Investigate the degreasing methods used in the machine shops with a view to designing an automatic system.

(b) Design the lay-out of a gold plating section and after organizing the installation of this plant control the section during the pre-production period.

6. Training in Drawing Offices

Students will vary in their ability and may need more or less practice in the basic principles. For instance, a man with a degree in physics rather than engineering may well need to spend a preliminary period in a drawing office training school. It is essential that during their period in the main drawing office all trainees should undertake at least one project on their own initiative. Mechanical problems are so closely involved with electrical ones that the trainee will generally appreciate the reason for training in this department. He must be given the opportunity to obtain as much knowledge of components, materials, tolerances and methods as time will allow. A typical example of such a project is:—"Given a standard type of telephone dial, design a completely sealed and self lubricating dial for field use." This sort of project might well be coupled with experience of methods engineering in which the trainee would be expected to progress the job to the finished state.

7. Training in the Laboratory

Training in a development laboratory could present some difficulties. Although the trainee is generally interested in this type of work it can still be a period of disappointment or boredom unless certain mistakes, which can easily be made, are avoided. An obvious difficulty is that the trainee spends only a short time in a particular laboratory during the second (or general) phase of his training. This clearly restricts the responsibility he can be given for any long term project. It is important that he should have experience of a variety of development work representative of the industry with which he is concerned, and also have the opportunity of working with a number of engineers so as to be able to benefit from more than one man's ideas on experimental techniques.

There is a danger that a hard-pressed departmental head could take the point of view that the man would not be long enough in the department to become conversant with the more important aspects of the work. He might, therefore, tend to employ the trainee on routine wiring and construction work without giving him an opportunity of making use of his technical knowledge. Here again the solution to the problem lies in the project system. If the department is in the habit of thinking of suitable items of laboratory work as projects for students then the problem is nearly solved. Care, however, is needed in the choice and supervision of the projects. The work of a laboratory is generally not of a repetitive nature, and no two projects will be exactly alike. There are, however, certain points which must be borne in mind.

Firstly, the trainee may find with some surprise that much of his theory does not quite work out in practice. The engineers he works with will seem to have an amazing amount of practical engineering knowledge and he must not be overawed by this. It may worry him that the approach to many development problems is empirical rather than theoretical, and it must be pointed out to him that theory is the basis for intelligent experimental work. The tasks that he is given must not be so involved that he cannot bring them to a satisfactory conclusion during the length of his stay in the department, and in particular he must not be left to struggle aimlessly with problems beyond the scope of his knowledge of experimental work. Many projects which start off in a simple fashion become complex at a later stage and require the guidance of a more experienced engineer. If this will not be readily available the project should not be used.

Some tests may call for the use of what appears to be a complicated mass of testing equipment, although the actual testing operations themselves are of a relatively simple nature. The engineer supervising the trainee must realize that what has become familiar apparatus to him may be quite confusing at an early stage of training. The requirements for the laboratory project thus evolve in the following way:

(a) It must be short enough to be brought to a reasonable stage of completion during the student's stay in the department.

(b) It should be based on some theory which he understands, and give experience of experimental techniques used in research and development.

(c) Help should be available at all stages, and where complex circuitry is involved the supervising engineer should at the outset break the circuit down into more familiar circuit configurations.

Most important of all, the student will learn from experience, but this must be experience with the guidance of established engineers always available. Much depends on the choice of the engineer who will perform this supervision. He must be willing to assist the trainee and to pass on his own knowledge freely, and the nomination of suitable people for this in the various departments is a matter of first importance.

Where appropriate there is something to be said for a brief course by a training instructor on the equipment developed in the laboratory group. The man's academic training may not have introduced him to laboratory techniques or specialized equipment. A short lecture course on these together with details of the systems being developed and the factors of cost involved could serve as valuable introduction to the laboratory period. References to relevant literature should also be provided. Examples of typical laboratory projects are given in Appendix 5.

8. Conclusion

It may be argued that an energetic trainee would find out all the necessary information without the project system and that this merely throws an extra load on the training staff and departmental supervision. The system however, has the following advantages:

(a) A report on specific details is required. Therefore, the trainee must make the effort to understand these points. Whatever the merits or shortcomings of his final report his attention will have been directed to the things that matter.

(b) To produce a satisfactory report the trainee must work at a similar rate to that required during his academic training. He will have to devote some time outside working hours to marshalling his thoughts and reading around the project subject. There will certainly be no time for boredom or inactivity.

(c) To obtain the information required he will have to develop a tactful approach to works personnel and realize that while he may lead the majority of them in theoretical knowledge, he lacks experience in practical matters.

(d) The department where the trainee is working has a fixed overall plan (though flexible in detail) and the temptation when hard pressed to side-track the trainee on to 'odd jobs' rather than organize his training is more readily resisted.

It is obvious that a considerable amount of work is required to draw up and revise the projects. It has, however, been found that students themselves can be of great assistance in analysing the basic work of the department and gathering information on which projects can be based—a useful project in itself! Time spent on this type of work is very worthwhile. Students who have followed both this and earlier types of training comment favourably about the new scheme. The content of their reports show that they are absorbing more information more quickly than was previously the case.

In addition one of the most encouraging features of the scheme has been the enthusiasm with which it has been received in works departments. Some supervisors were found on their own initiative to be applying a modified form of the scheme to grades other than professional trainees with apparent success.

This approach sees the end of much of the reluctance to pass on information and supervisors are pleased to see from the final report that their training efforts have been successful. Operators who hitherto had shown little interest beyond their own particular job are often stimulated into asking questions and gaining further knowledge. A careful follow-up of all projects by the training department remains essential.

The technique described is merely an approach to a very big problem, but it does possess the advantage of having the flexibility required of any scheme of professional training in industry.

9. References

1. "The Training of Graduates", The Institution of Electrical Engineers, London, June 1960.
2. D. L. Marples, J. F. A. Radford and J. L. Reddaway, "An approach to the techniques of graduate training", *Proc. Instn Mech. Engrs*, **170**, No. 22, p. 747, 1956.
3. "The Practical Training of Professional Radio and Electronic Engineers"—A report prepared by the Education and Training Committee of the British Institution of Radio Engineers, *J. Brit.I.R.E.*, **23**, No. 3, pp. 171–6, March 1962.
4. G. M. Russell, "Group project training for graduates", *Chartered Mechanical Engineer, J. Instn Mech. Engrs*, **8**, No. 7, p. 430, September 1961.

10. Appendices

10.1. *Example for a Machine Shop*

This concerns a large production machine shop manufacturing a wide variety of products. The main object of the project is to give the trainee an understanding of the uses and limitations of hand and machine tools and to give an appreciation of the materials and techniques used. He must also be given a certain amount of practical experience in the use of these tools. A further objective is to provide experience of the organizational problems of the department and to enable the trainee to meet and mix with people at both shop floor and supervisory level, an end never achieved in a training school. The project thus falls into two parts. In practice the order in which the two parts of the project are undertaken is not important, and there are arguments for and against the practical part being first. In any event if the training facilities are to be used to the full it may be necessary for one group of trainees to attempt one part of the project while another group follows the other part.

To complete the picture of a production unit the project also includes the associated finishing shop, heat treatment department, mechanical inspection department and production control unit.

A supervisor is nominated in each department to give the trainee reliable assistance and information in addition to that obtained from the operatives. This person's name is clearly shown on the project sheet for the section concerned. The following is an extract from a project sheet as issued to a trainee.

GRADUATE AND STUDENT TRAINING PROJECT

Machine and Finishing Shops, Mechanical Inspection Department and Production Control

During the project you will spend periods in the following:

Machine Shops Milling
Power presses and guillotine
Capstans
Switch machine shop
Drilling and tapping
Automatics
Dial machine shop
Heading and thread rolling

Finishing Shops Buffing and polishing
Dipping and plating
Spraying and painting

Mechanical Inspection Department

Heat Treatment Section

Production Control

Period A (four weeks)

During this period you will gain practical experience of machine shop work. You will be given drawings and lay-out cards for test pieces that require milling, turning and drilling operations, and the use of hand tools. You will follow the normal machine shop procedure in drawing raw material from stores, securing tools, and having your work finished and inspected.

Some experience of actual production work will be given. At the start of your training in the machine shops you will be given booklets on the safety requirements for the various machines. Study them well. The Safety Officer will give you practical instruction on safety points.

Period B (four weeks)

During these four weeks you are required to study in general terms, the following aspects of the work done in the machine shops and connected departments.

(1) The type of work done and the materials, machines and labour employed.

(2) The organization of the shops and the allocation of work throughout them.

(3) The procedures followed in the shops from the initial receipt of a work order to the final inspection and despatch to store.

Apart from the specialized work of the departments that you visit you should take special note of such general problems as:

Precautions taken for the safety of persons and equipment.

Transport of goods on and between sections.

Physical lay-out of the shops.

Provision of specialized shops for switch and dial machining.

Payment of operators, i.e. piecework, time rate and bonus schemes.

Relative efficiency of part, full and overtime working.

A report will be required at the end of your training in the department and should include not only answers and comments on the topics raised here, but any personal comments, criticisms and suggestions or improvements that have occurred to you during your stay. The report will be read by both members of the machine shop supervision and training department.

Typical examples of 'pointers' given for separate sections.

(The next section of the project sheet gives 'pointers' for separate sections. Because of pressure on space only two examples are detailed below.)

Power Presses and Guillotines (one day)

Types and sizes of presses and guillotines.

Automatic presses.

Type of work done and speed of operation etc.

Advantages and disadvantages.

Operations including blanking, piercing, bending, drawing, shearing, etc.

Tools. 'Blank and follow on' tools, 'compound' tools.

Machine guards.

Limitations of belt driven machinery.

Do you think the work is suitable for female operators?

The lay-out of the shop, lighting, air-conditioning, etc.

Types and quantities of raw materials used.

What happens to the blanked out scrap?

Why is some material bought cut to size while other material has to be guillotined?

Inspection Department (three days)

(a) The main purpose of the inspection department is to prevent faulty pieceparts *being made* and the place where this must be done is at the point of production, *the machine*. Are adequate steps being taken to ensure this?

(b) 100% inspection. Is it always economical?

(c) Inspection of outside purchases. Are we doing another firm's job?

(d) Allocation of scrap charges. Do the quantities of scrap seem excessive? What practical steps could be taken to prevent excess scrap?

(e) The 'tools of inspection'. All types of gauges, etc. The vernier scale. What are the main advantages of the comparator over gauges in, for example, the inspection of screw threads? Could it be more widely used? What is implied by "statistical quality control"?

(f) Can any piecework or bonus scheme be successfully applied to inspection personnel?

(It will be noted that in each case, in addition to pointers to the specialized work of the section, one point of general production interest is raised as well.)

10.2. *Example for a Mass Production Assembly Shop*

The opportunities for planning production and for supervising operators are not easily available here, but there are other factors resulting from the mass production itself that form the basis for useful project subjects. As always, an attempt is made to provide an opportunity for the trainee to gain some practical experience of the work involved, but it must be so introduced as not to interfere with the production of the department.

An extract from a typical project is shown below:

Graduate and Student Training

Television Assembly

You are required to study and report on the work of the department by means of the following exercises:

(1) Assemble a complete television receiver in the following stages. Drawings, stock lists and components are available from the supervision in the sub-sections concerned.

(a) Assemble the printed wiring boards used in the set from the drawings and stock lists. The complete assembly should be taken to Mr. ...... for inspection and then taken for dip soldering. You should then follow the panel through the test procedure; for this refer to Mr. ....

(b) Obtain from the supervisor a cable-form board together with drawings and construct a cable form.

(c) Assemble the main chassis of the television set using the sub-assemblies already made.

The completed chassis should be taken to Mr. ...... for inspection and you should then follow it through the test procedure. (Approximately 4–5 days.)

(2) Draw a 'flow chart' covering the production of a television receiver. Individual repetitive operations need not be specified, but may be grouped together as a single operation, e.g. "Assemble ten components to panel" could be considered as one operation provided that no major transport was involved.

Is the method of assembly 'fluid' enough to allow new designs of set to be introduced on the same production lines in future years?

Consider in what respect the use of an intermittently moving conveyor belt is an advantage over the methods of

(a) No moving belt.

(b) A moving belt from which the work is extracted, the job done and the work replaced on the belt.

(c) A slow continuously-moving conveyor.

A conveyor system of work movement is considered by some people to be in the interests of production only, and not in the interest of the personnel performing the operations. Is this in fact the case? Why is so much use made of compressed air as a prime mover in this department? Consider other prime movers and explain why these are not used to any great extent.

There are six major inspection points involved in the production of the television set. Visit these points in turn and analyse all rejects to find the most frequent cause of failure. In your opinion is it necessary to have two inspection points on the main assembly track, or would it be better to have only one 'reject loop' at the end of the main assembly track? (Approximate time three days.)

(3) *Material Handling*

Detail the method of transport for the various components and sub-assemblies. Some items are delivered daily and others weekly. Consider the relative advantages of these systems, and suggest improvements particularly where the condition of the items is affected during the transport.

(4) *Process Control*

Discuss the control of quantities of items used in the department, e.g. cabinets per hour, cable forms per hour. How is the demand met and excessive stock-piling avoided? What would be the effect of the supply of any items being insufficient?

(5) *Safety*

What methods are adopted to comply with safety regulations? Consider any means by which the safety factor could be increased and discuss these with the supervisory staff.

(6) Write a report on the whole of the project including any criticisms, on the form which the project takes.

(Two days are allowed for Sections 3, 4, 5 and 6.)

10.3. *Example for a Test Department*

The following is an extract from a typical project sheet issued to the trainee.

Transmission Test and Assembly

The section is responsible for testing all the sub-assemblies used in transmission systems, e.g. channel panels, channel amplifiers, etc., involving three hundred to four hundred different types of panels.

Details of the Project

(i) Follow the programme detailed below and familiarize yourself with testing periods.

(ii) Complete the assignments allocated by the supervisor.

You must carry out full test procedures on the equipments in the department. The following points should be included in your final report.

(a) The paper work involved in 'booking in' and 'booking out' panels and the changes in the routine for panels for urgent jobs. References should be made to Rejects, Production Control, day and time stamping, and three monthly schedule, nomination cards and rack targets.

(b) Discuss the uses of a channel panel with reference to its applications in a carrier system, in the voice-frequency band, and in out-of-band signalling and telegraph working.

(c) Describe the hybrid transformer and its working with aid of drawings.

How is carrier leak suppressed?

Describe the operation of an oscillator and static relay.

How is limiting achieved?

Why is a high pass filter included in the circuit?

What is the principle of a harmonic generator?

At the end of the report give your personal comments on the syllabus together with any suggestions for improvements.

Details of Programme

Type 51 and 56 Equipment

Channels, amplifiers, harmonic generators and modulators are basic pieces of transmission equipment and you will see all of them tested.

(a) Because of numbers the panel testing has been split into four rigs each doing three or four tests. Spend one day carrying out the tests on each rig. Follow the tests on a schematic drawing of the equipment.

(b) *Oscillators (one day)*. Two types of oscillator are in general production. Note the investigation of inversion, the variation of temperature and voltage and the ageing tests.

(c) *Amplifiers (one day)*. Explain the operation of the automatic testing apparatus.

Bibliography

T.T.I. Sheets.

G.E.C. Systems Handbook.

B.S.S. Terms and Symbols in Telecommunications.

10.4. *Example for a Production Control Department*

Extract from a project sheet for a small specialized production control section

Production Control—Relay Section (one week)

(a) Draw a flow chart illustrating the system by which production of relays is controlled after an indent is received. N.B. It is not necessary to include every detail of the paper-work involved.

(b) What are the factors determining the size and content of a 'batch'? Explain why the batch system is considered necessary.

(c) What is the purpose of the 'A' file?

(d) Investigate the procedure involved in terms of relevant paper work and action, if for any reason a relay is not produced during the week for which it is programmed. List some possible reasons for the production falling behind schedule.

(e) What steps are taken to avoid over-production of more specialized parts such as spring-sets, whereas coils are slightly over-produced in the normal way?

(f) Consider the lay-out, use and position of the channel panel racks in order to determine whether improvements are possible.

(g) Discuss the use of photo-copying machines in the department.

(h) Criticize the present control system. Suggest an alternative scheme.

General

At the end of the training period prepare a report giving your personal observations on the above syllabus together with any suggestions for improvement. The report will be discussed by the Training Officer and the Supervisor of the department concerned.

10.5. *Example for a Development Laboratory*

Examples of laboratory projects are as follows:

(a) The development of a transistorized channel panel capable of working at a higher temperature than the existing equipment.

(b) The development of transistor divider circuits.

(c) Modification of the lay-out of various electronic bricks to reduce the number of components and the overall space required.

(d) To design a low-pass filter to meet the given specification. This involves winding various coils and the use of desk calculator.

(e) The development of power failure alarms for signalling equipment.

(f) The design of a miniature transistor master oscillator.

Manuscript first received by the Institution on 20th November 1962 and in final form on 8th April 1963. (Paper No. 839/Ed 11.)

U.D.C. 621.396.967.2

An Integrated Marine Radar System

By

A. HARRISON, B.Sc.
(*Associate Member*)†

AND

D. CHAMBERLAIN†

Presented at a meeting of the Radar and Navigational Aids Group in London on 9th May 1962.

Summary: Fault records on a particular marine navigational radar are analysed to determine the probability of successful performance over a given interval. This indicates that certain units are much less prone to failure than others. The duplication of equipment to give greater reliability is examined, and it is shown that from the reliability point of view it is more efficient to duplicate only the least reliable units, and to provide facilities so that any combination of the available duplicates may be switched into use in the complete system. A comparison is drawn between two installations using the same units, one consisting of two complete but independent radar sets, the other providing higher reliability by the use of fewer units with switching facilities.

Several switching arrangements are in seagoing use, and examples of these are given, showing that in addition to increased reliability they provide operational facilities not available on a single equipment. Some of the problems overcome in the design are discussed, particularly the interaction of human engineering and switch mechanisms in a multiple switching operation, and the influence of these factors on the final design.

S.S. *Canberra* carries the most complex equipment of this type to date, and the paper concludes with an illustrated description of the installation.

PART 1

Statistical Approach to Fault Analysis in System Design

1. Introduction

The systems engineer is accustomed to using fault records of equipment in operational use to indicate where modifications to current production will improve reliability. This paper deals with another aspect of the use of fault records, which is perhaps not so well known, namely, to predict the behaviour of proposed new arrangements of existing units, and to provide a sound basis on which to choose one system in preference to another.

This analysis shows how to make the best use of what is available; it leads logically to the arrangement of units which makes up the radar equipment in S.S. *Canberra*, and it provides a reference against which the reliability of the equipment in operation can be judged.

The electronics engineer is continually facing a demand for higher reliability, and he attempts to meet this by using improved components and units.

When everything possible has been done to a piece of equipment in this direction but the demand is for an even higher reliability, the only approach is to duplicate the equipment. Ships' compasses, chronometers in pre-radio days, and the duplication and even triplication of units of an aircraft auto-pilot are examples of this principle. When a ship is proceeding in fog in the presence of other ships, or is committed to a narrow channel, the result of radar failure could be catastrophic. Hence a need has arisen for duplication of radar equipment on board ship.

The obvious approach is to fit a second complete radar set, but this is not necessarily the best solution. Considerable cost saving can be obtained as well as a higher improvement in reliability by a different arrangement. To see how this can be achieved we must consider the faults which occur in equipment.

Fault records will provide information on the total number of faults in the time interval chosen. In order to draw trustworthy conclusions the test period must extend over a long time, and the records must include a large number of faults, occurring on many installations. Having obtained the test figures, we encounter the first problem. If 50 per cent of the sets failed in 50 000 working hours, can we say that 10 per cent would fail each 10 000 hours? Obviously not.

† Kelvin Hughes Division of S. Smith & Sons (England) Ltd., Dagenham, Essex.

It is common experience that after the initial fault period the rate of failure is low, and increases to a maximum by the time half the sets have failed. After this the rate decreases again, since some sets obstinately refuse to fail and like old soldiers, never die but only fade away. Hence the time intervals chosen will have considerable effect on the conclusions to be reached.

The first time interval during which the data are collected must be sufficiently long to indicate the maximum rate of failure. The second interval must bear a sensible relationship to the periods of operational use of the equipment, and the maintenance facilities. How then are we to relate the figures for the two intervals?

Francis Bacon could well have been thinking of this difficulty when he wrote: "If a man will begin with certainties he shall end in doubts, but if he will be content to begin with doubts he shall end in certainties." We follow this advice by considering not the number of faults in a given time, but the *probability* that the equipment will continue to operate throughout a chosen time interval, or will suffer one or more faults in that time. The mathematical frame on which the argument is erected is well expressed by Moroney in his book "Facts from Figures",† where he shows how to deal with the probability of a number of isolated events occurring in a given time.

The term "isolated events" has a particular meaning here. Suppose previous records show that a proportion A of the articles manufactured is defective and the remaining proportion B is good ($A+B = 1$). If at a later date we take a sample of m articles, and the process is unchanged, then one would expect a certain number of the m articles in the sample to be defective. The probability of any number of articles between 0 and m being defective is given by the successive terms of the binomial expansion of $(A+B)^m$. In particular, the first term of the expansion gives the probability of the sample containing no defective articles.

The number of faults which occurred may be known, but the number of those faults which did *not* occur, may not be known, so A, B, and m cannot be assigned values. One-sided information of this type is often met—the number of dust particles in a litre of air, the number of industrial accidents per year, and so on. Occurrences of this type are defined as isolated events. The binomial expansion cannot be used to forecast their probability density, but fortunately certain characteristics of the data allow it to be manipulated to a form which enables us to use a different expansion.

† M. J. Moroney, "Facts from Figures", (Penguin Books, London, 1951).

2. Poisson Distribution

In this case the service fault data reveal a large number of faults n, occurring at random during a large time interval T, so that the average fault density λ is n/T.

Now consider an interval t, small compared to T. The probability that any particular single fault occurs within the small interval t is t/T, i.e. $p = t/T$. The probability that it does not fall in the interval t is $q = 1-p$. Since the fault density is λ the average number of faults in the interval t is $t\lambda$. Calling this average z, then $z = t\lambda = tn/T = np$.

The probability P_x that x faults will occur in the time interval t is given by the $(x+1)$th term of the binomial expansion $(p+q)n$ and since $q = 1-p$ we may write

$$P_x = \frac{n!}{x!(n-x)!} \cdot p^x \cdot (1-p)^{n-x}$$

$$= \frac{n!}{x!(n-x)!} \cdot \frac{1}{n^x}(np)^x \cdot \left(1 - \frac{np}{n}\right)^{n-x}$$

$$= \frac{1}{x!} \cdot \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots$$

$$\cdots \left(1 - \frac{x-1}{n}\right) \cdot z^x \left(1 - \frac{z}{n}\right)^{n-x}$$

Since n is large, the factors $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)$ etc., approximate to 1, and $\left(1-\frac{z}{n}\right)^{n-x}$ approximates to e^{-z}.

Hence

$$P_x = \frac{z^x e^{-z}}{x!}$$

Thus if we know z, the average number of faults in a relatively small time interval t, we can calculate the probability P_x of the occurrence of any chosen number of faults x in that interval. The successive terms are:

$x = 0$: probability of 0 faults $= P_0 = e^{-z}$

$x = 1$: probability of 1 fault $= P_1 = ze^{-z}$

$x = 2$: probability of 2 faults $= P_2 = \frac{z^2}{2!} e^{-z}$

$x = 3$: probability of 3 faults $= P_3 = \frac{z^3}{3!} e^{-z}$

and so on.

This distribution of probabilities is called the Poisson Distribution and is expressed graphically in Fig. 1 for three typical values of z.

For the purpose of the equipment in mind there are two probabilities; the first is P_0, the probability

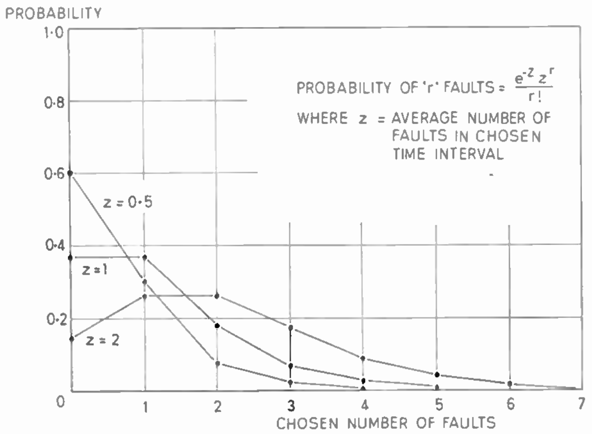


Fig. 1. Poisson Distribution.

of no faults, i.e. the probability of the equipment working successfully throughout a time interval; the second is the probability of the equipment failing due to one or more faults, which is the sum of the successive probabilities $P_1+P_2+P_3+\ldots$, which is of course equal to $1-P_0$. (Probabilities are usually expressed as decimal fractions.)

The notation is simplified by saying that the probability of the equipment working with no faults P_0 is equal to S, the probability of success.

The probability of equipment failing is

$$P_1+P_2+P_3+\ldots = 1-S = F,$$

the probability of failure.

2.1. *Combination of Unit Probabilities*

In combining various arrangements of units to make complete installations, unit probabilities must be combined to obtain installation probabilities. Figure 2 shows the two basic configurations.

Where a number of units are combined in a chain to make up a complete equipment, so that the failure of any one unit puts the equipment out of action, the probability of the equipment working is given by the product of the probabilities of success of the individual units. For example, if $S_1 = S_2 = S_3 = 0{\cdot}5$ or 50%, the probability of the chain of three units working is $0{\cdot}5\times0{\cdot}5\times0{\cdot}5 = 0{\cdot}125$ or $12\frac{1}{2}\%$.

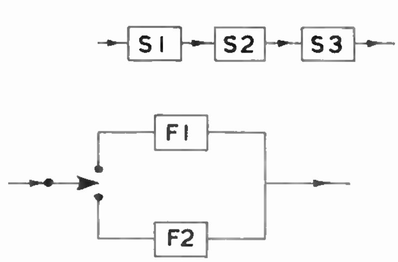


Fig. 2. Basic unit configurations showing success and failure probabilities.

Units may be connected into an equipment by switches, so that in the event of failure of one unit, an alternative similar unit may be brought in to service. In this case the probability of failure of both units at the same time is given by the product of the probabilities of failure of each unit.

Thus if $S_1=S_2=0{\cdot}7$, then $F_1=F_2=1-0{\cdot}7=0{\cdot}3$.

Probability of failure of both units

$$= 0{\cdot}3\times0{\cdot}3 = 0{\cdot}09$$

Probability of success of both units

$$= 1-F = 1-0{\cdot}09$$
$$= 0{\cdot}91 \text{ or } 91\%.$$

This may now be combined with the probability of other units working by multiplying as above, to give the probability of the whole equipment working.

3. Fault Report

A survey of 40 ships was made during the first six months of 1961. The population of equipment being considered was installed on ships and was selected from a production period covering the six months prior to the trial, which itself must cover six months to provide enough operational hours. In this time of course, it is likely that modifications will have been introduced to deal with any obvious troubles, and this must be taken into account in considering this analysis and any action it suggests.

The fault data have been summarized as follows:

Operational hours (to nearest 100 hours)

24 ships each working 600 hours average in 6 months	14 400 hours
16 trawlers each working 2100 hours average in 6 months	33 600 hours
Total	48 000 hours

Table 1

Faults on 40 ships over six months

Unit	All Faults	Major Faults
Transmitter	68	18
Display and power unit	61	48
Generator	2	0
Aerial	7	3
Complete equipment	138	69
Average life between faults	350 hours	700 hours

3.1. *Fault Classification*

The faults occurring were analysed in detail, but for this purpose they have been classified as

(a) Minor—which could be corrected using spares

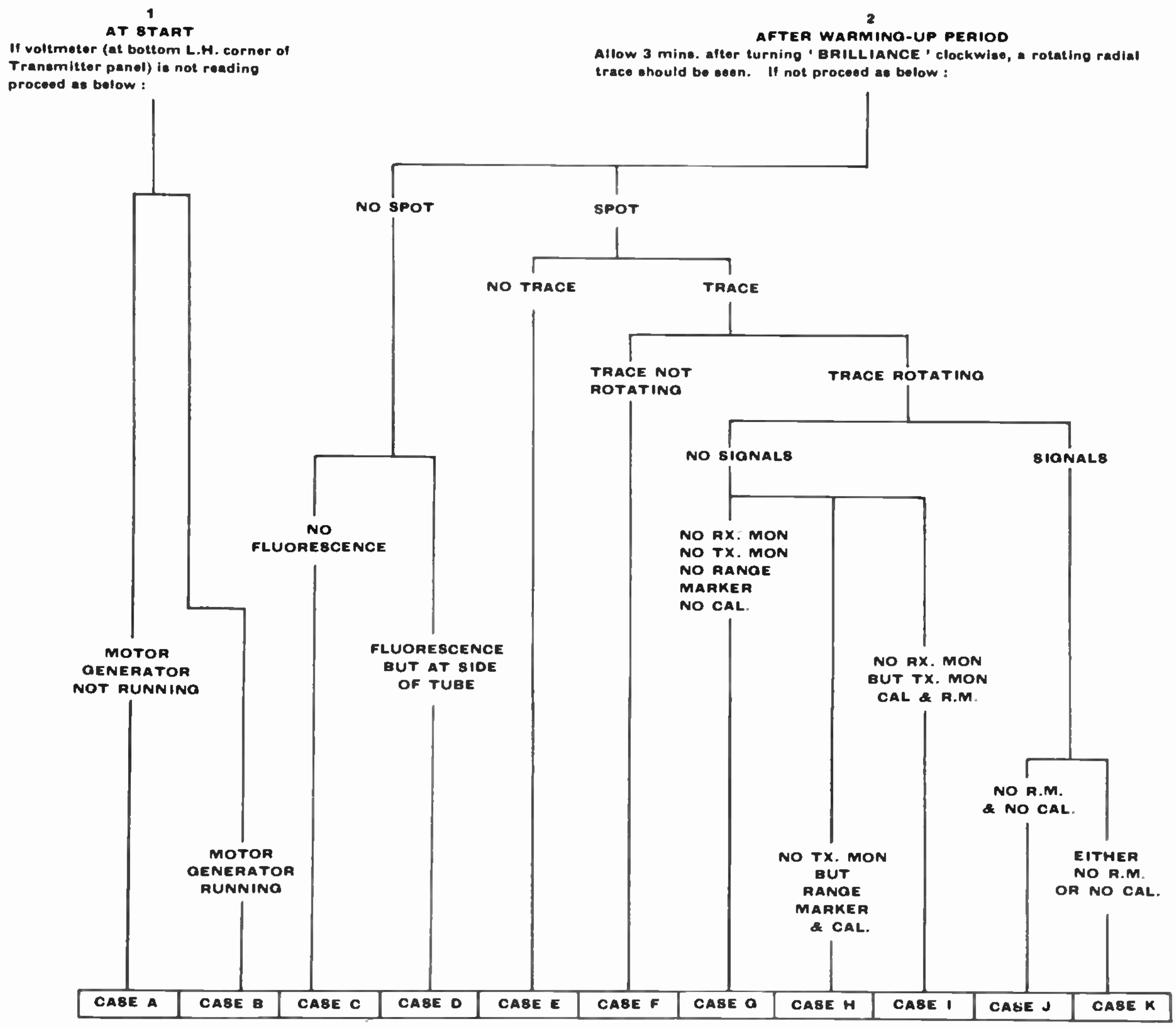


Fig. 3. Typical fault location chart.

and tools available on board ship (e.g. valve replacement)

(b) Major—which required depot services to restore the set to use.

3.2. *Time Intervals*

A major fault cannot be corrected until depot services are available. This might be say 300 working hours, and this time has been used for calculations on the probability of major faults. Minor faults are omitted in this period, since several could occur, and could be corrected, without appreciable effect on the figures.

If one unit of a dual system is out of action due to a minor fault, it would seem relevant to consider the probability of a second fault of either kind occurring in the other units, during the time taken to repair the first fault. Assuming this time is three hours, we must consider the probabilities of all faults in this interval.

Three hours may seem a long time in which to repair a minor fault, but the man dealing with the fault is a trained navigator, not an electronics engineer. With the assistance of a chart like the one shown in Fig. 3, he is assisted to diagnose the area of trouble, and can follow the action suggested in the handbook. In nearly all cases he can get the set operational again, but his progress is slow and steady. Experience shows that a time of three hours is a fair average.

4. Calculation of Probabilities

4.1. *Three hour Interval*

Taking all faults on one complete set of equipment into consideration, the average number of faults in three hours

$$= z = \frac{138}{48000} \times 3 = 0{\cdot}00862$$

$$e^{-z} = 0{\cdot}9914$$

Probability of 0 faults in 3 hours

$$= e^{-z} = 0{\cdot}9914\ (99\%)$$

Probability of 1 fault in 3 hours

$$= ze^{-z} = 0{\cdot}00854\ (1\%)$$

Probability of 2 faults in 3 hours

$$= \frac{z^2}{2!}e^{-z} = 0{\cdot}00037\ (0\%)$$

and so on.

Thus $S = 99\%$ $F = 1\%$.

4.2. 300 *Hours Interval*

Taking major faults only on one complete set of equipment into consideration, the average number of faults in 300 hours $= z = 0{\cdot}431$.

$$e^{-z} = 0{\cdot}6499$$

Probability of 0 faults in 300 hours

$$= e^{-z} = 0{\cdot}6499\ (65\%)$$

Probability of 1 fault in 300 hours

$$= ze^{-z} = 0{\cdot}2807\ (28\%)$$

Probability of 2 faults in 300 hours

$$= \frac{z^2}{2!}e^{-z} = 0{\cdot}0604\ (6\%)$$

Probability of 3 faults in 300 hours

$$= \frac{z^3}{3!}e^{-z} = 0{\cdot}0084\ (1\%)$$

and so on.

In round figures, in a 300 hours interval a set is 65% likely to run without fault, and 35% likely to fail. To obtain a clearer impression of what this means in actual operation, it should be remembered that 300 hours represents a trawler trip, or three months use on a deep-sea ship.

4.3. *Unit Failure Probabilities*

4.3.1. Major faults in a 300 hours interval

Table 2 shows an analysis of major faults which are probable for a 300 hours interval.

Table 2

Unit	Faults	z	$e^{-z}=S$	$1-S=F$
Transmitter	18	0·1125	0·8936	0·1064
Display and power unit	48	0·3000	0·7408	0·2592
Generator	0	0	1·0000	0
Aerial	3	0·0187	0·9814	0·0186
Equipment	69	0·4310	0·6499	0·3501

There are two points to note. Firstly, the generator and aerial are at least one order better than the display and transmitter in failure probability.

Secondly, it was previously stated that

$$S = S_1 \times S_2 \times S_3 \text{ etc.}$$

This may now be checked by multiplying the figures above in the S column. On a slide rule this gives 0·650 (S for complete equipment is calculated as 0·6499 in Section 4.2).

4.3.2. All faults in a 3 hour interval

Table 3 shows an analysis of all faults which are probable for a 3 hours interval.

Table 3

Unit	Faults	z	$e^{-z}=S$	$1-S=F$
Transmitter	68	0·004250	0·99577	0·00423
Display and power unit	61	0·003812	0·99623	0·00377
Generator	2	0·000125	0·99987	0·00013
Aerial	7	0·000437	0·99956	0·00044
Equipment	138	0·008625	0·99137	0·00863

Again the difference in failure probability for the generator and aerial is about an order of magnitude better than the display and transmitter.

In the search for higher reliability by duplicating equipment, the figures suggest that there is little point in duplicating the aerial and the generator. It will be shown that duplication of the transmitter and the display improves their unit reliability to the same order as that of the aerial and generator, giving a well-balanced design in which any unit is no more liable to fail than any other.

5. Comparison of Two Systems

5.1. *Details of the Two Systems*

The obvious way of duplicating the equipment is to fit two independent radar sets, each consisting of an aerial, a transmitter, a display, and a generator, and this has been done on a number of ships. This is illustrated in Fig. 4(*a*). The alternative is a switchable dual set, consisting of one aerial, two transmitters, two displays and one generator (Fig. 4(*b*)). The switch boxes needed for this are considered to be at least as reliable as a generator and are omitted from the calculations. This is borne out by the detailed analysis of the original fault data, in which not one switch failed on 40 ships in six months, and also by experience of a final test of sets in the factory.

The probabilities of success S and failure F are computed as before.

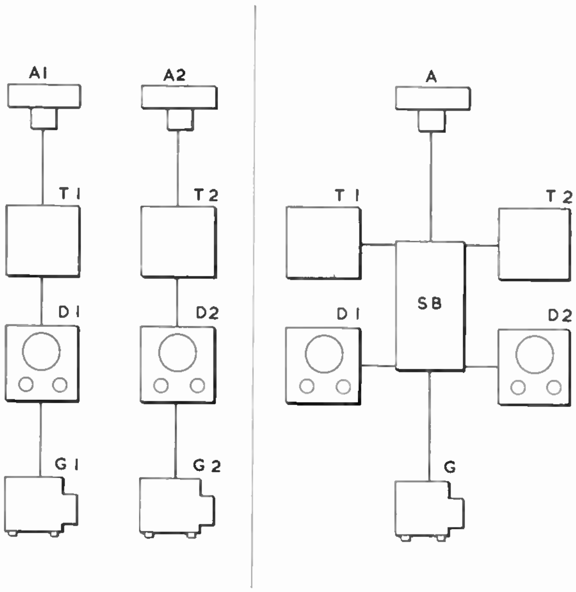


Fig. 4. Basic radar systems.
(*a*) Two independent sets. (*b*) Switchable dual set.

First, for alternative units,

Probability of failure is $F = F_1F_2 = F_1^2$ for identical units

Probability of success of switched unit $S = 1 - F_1^2$

Second, for complete equipment,

$S = S_1 \times S_2 \times S_3 \times S_4$.

5.1.1. Major faults in a 300 hours interval for two independent sets

$F = F_1^2 \quad = 0{\cdot}35^2 \quad = 0{\cdot}122$ or 12%.

Therefore

$S = 1 - F = 0{\cdot}878$ or 88%

Using a duplicated transmitter and display,

Transmitter

$F_1 = 0{\cdot}1064 \quad F_1^2 = 0{\cdot}0113 \quad S = 0{\cdot}989$,

Display

$F_1 = 0{\cdot}2592 \quad F_1^2 = 0{\cdot}0673 \quad S = 0{\cdot}933$

for a single common aerial and generator

Aerial $S = 0{\cdot}981$,

Generator $S = 1{\cdot}000$.

For complete dual equipment

$S = 0{\cdot}905$ or $90{\cdot}5$%, and

$F = 0{\cdot}095$ or $9{\cdot}5$%,

showing a small balance in favour of switching.

Note that in the final column of figures, the values of S are now all of the same order. This verifies the statement that duplication of the transmitter and display produces a well-balanced design from this point of view. The important point, however, is that because the switch facilities allow four possible arrangements of the available units, the probability of failure of the whole installation is reduced from 12% to $9\frac{1}{2}$%.

5.1.2. All faults in a 3 hours interval for two independent sets

$F = F_1^2 = 0{\cdot}000074$ or $0{\cdot}0074$%

$S = 1 - F = 0{\cdot}999926$ or $99{\cdot}9926$%

This sort of figure is not unexpected. The chance of two sets failing in the same three hour interval must be exceedingly small when each has an average life of 350 hours between faults. From the point of view of comparing the reliability of two systems, however, this figure is useless, since its significance does not really extend beyond two or three digits.

There is obviously no point in computing the figures for the alternative systems, since they will not differ by any significant amount. On the original basis of comparison, we cannot distinguish between the reliability of the two systems.

The basis of comparison is therefore changed, and the relative reliability of the two systems in the presence of one existing fault, which is placed arbitrarily in one transmitter, will be considered.

5.2. *Existing Fault in One Transmitter*

5.2.1. Major faults in a 300 hours interval

(a) Two independent sets

$S = 65\% \quad F = 35\%$.

(b) Switchable dual set

Display $F = F_1^2 = 0{\cdot}0672$	$S = 0{\cdot}933$
Remaining transmitter	$S = 0{\cdot}8936$
Aerial	$S = 0{\cdot}9814$
Generator	$S = 1{\cdot}0000$

The product of these figures, is $S = 0{\cdot}821$ for the switchable system from which $F = 0{\cdot}179$ or 18%, i.e. half the failure probability calculated for the independent system.

5.2.2. All faults in a 3 hours interval

(a) Two independent sets

$S = 99\% \quad F = 1\%$

(b) Switchable dual set

Display $F = F_1^2 = 0{\cdot}0000142$	$S = 0{\cdot}99998$
Remaining transmitter	$S = 0{\cdot}99577$
Generator	$S = 0{\cdot}99987$
Aerial	$S = 0{\cdot}99956$

The product of these figures is $S = 0{\cdot}9952$ from which $F = 0{\cdot}0048$, again about half the failure probability of the independent system.

5.3. *Summary of Probabilities*

Table 4 summarizes the figures obtained. The last two columns indicate that, as compared with a single radar set, the chance of losing radar information at a critical moment is reduced by a factor of 3 for the time considered, by duplicating the whole equipment. The alternative switched arrangement halves this risk again, to give a safety factor of 6 over a single equipment.

Table 4

Equipment	F	S	F†	S†
300 hours Major faults				
Single	35%	65%	100%	0
Independent	12%	88%	35%	65%
Switched	9·5%	90·5%	18%	82%
3 hours All faults				
Single	1%	99%	100%	0
Independent	—	99·9%	0·9%	99·1%
Switched	—	—	0·5%	99·5%

† In these two columns one fault is assumed to exist in one transmitter.

5.4. *Note on Cost*

The financial aspect is also interesting. Duplication of the whole equipment of course doubles the capital cost of the installation.

The switched arrangement described saves the capital and installation costs of one scanner and one generator, and adds the cost of one switch-box. This gives a net saving of about 15% of the capital cost of one installation and also reduces installation problems. This could well mean a saving of 25% on the second installation.

6. Development of the System used in S.S. *Canberra*

6.1. *Operational Requirements*

Section 5.2 indicates the next logical step leading to the installation on S.S. *Canberra*. The 'success' figures are of the same order, but the display has the lowest probability of success. Hence, if we are adding one further unit to improve reliability, it should be a third display.

This has the additional advantage that it improves the operational facilities.

Considering the major faults in a 300 hours interval (Table 5), this calculation shows the expected effect which is to reduce the probability of failure from 9·5% to 4·7%.

Table 5

Two transmitters	$F^2 = 0{\cdot}0113$	$S = 0{\cdot}989$
Three displays	$F^3 = 0{\cdot}0174$	$S = 0{\cdot}983$
One aerial		$S = 0{\cdot}981$
One generator		$S = 1{\cdot}000$
For the complete installation	$S = 0{\cdot}953$	
	$F = 0{\cdot}047$ or 4·7%	

It is not worthwhile to carry the mathematical analysis much further, since the junction box reliability should be introduced. In addition some units, like the photographic projector display, are outside the scope of the original fault records.

There are, however, other aspects of a ship installation which were considered in the final design of the S.S. *Canberra* equipment. There are operational advantages in being able to use one transmitter for long range navigation, using long pulses and a narrow-bandwidth receiver to make the most of small signals, while the other transmitter is operating with short pulses and a wide-band receiver for high definition at shorter ranges. Besides requiring two scanners, this introduces the problems of the cross-interference of the systems, and some power supply problems.

The modulators used in the transmitters are synchronous—that is, one r.f. pulse is transmitted for each cycle of the supply. To minimize the effect of interference, the pulses are interlaced by operating the transmitters on opposite phases of the same supply. The interference on the display associated with one set from the transmitter of another then appears as a ring at about 35 miles, and not a 'snow storm' of pulses scattered at random over the screen.

The power required for the simultaneous operation of two radar sets would be too much for the standard generator. Two generators were therefore coupled and run from a common motor. In view of the novelty of this arrangement, it was thought desirable to duplicate it.

6.2. *Final Installation*

The final arrangement of equipment for the S.S. *Canberra* was therefore as follows:

Two 10-ft Scanners. These were mounted side by side on a platform on the foremast, giving a good all-round view with no obstructions.

Two Transmitters. Standard type-14 units.

Three Displays. (a) 16-in. c.r.t. true plot display; (b) 12-in. c.r.t. relative or north stabilized display; (c) 30-in. photographic bright projection display, with true plot facilities.

Two double generator sets and their control gear with a change-over switch-box.

Transmitter and aerial selector switch-box, permitting twelve modes of operation of the equipment.

Slave selector and indicator panel. This was necessary, to show the Officer of the Watch the mode of operation of the equipment in use. By a system of coloured lights, it might indicate for example, that transmitter 1 was coupled to aerial 1 and was operating under the control of the projection display, with the 12-inch display as a 'slave' to this. Transmitter 2 and aerial 2 might be operating under the control of the 16-inch display, and the whole equipment was being powered by generator set No. 2. The possible configurations are discussed in more detail in Part 2.

6.3. *Probability of Failure*

Considering major faults in a 300 hours interval the probability of failure is as shown in Table 6.

Table 6

Two transmitters	$F^2 = 0{\cdot}0113$	$S = 0{\cdot}989$
Three displays	$F^3 = 0{\cdot}0174$	$S = 0{\cdot}983$
Two aerials	$F^2 = 0{\cdot}0004$	$S = 0{\cdot}999$
Two generators	$F^2 = 0$	$S = 1{\cdot}000$
For the complete installation	$S = 0{\cdot}970$	
	$F = 0{\cdot}03$	or 3%

Because of the reasons given—changes of some units and so on—it is perhaps academic to compute the reliability of the equipment from the initial data, but the calculation does indicate a 3% probability of failure in 300 hours. It must be remembered that operational requirements as well as reliability, determined the duplication of some units.

6.4. *Operational Experience*

The equipment has now operated over 3000 hours. The previous data indicate that in this interval about one scanner fault, 5 transmitter faults, 12 display faults, and a fraction of a generator fault—say a total of 18—could be expected and the reports have confirmed this prediction.

During this period there were, in fact, 8 major faults and 13 minor faults. The older units—transmitter, cathode-ray tube displays and the generator—are responsible for one major and 5 minor faults only, indicating that the primary use of fault reports, which is, of course, to feed back corrective modifications to production, has had the desired effect of improving unit reliability. The newer units—10-ft scanners, 30-in. bright projection display and the switch-box account for 7 major and 8 minor faults. Five of the 7 major faults were repetitions—the finish on a shaft was incorrect, causing failure of an oil seal, and until the unit could be replaced, the only action was to replace the seal. Similarly, 3 minor faults were repetitions which have now been corrected. A long period of relatively fault-free operation may therefore be expected.

Two units have operated the whole time to date without a single fault, namely, the 12-inch display and the switch-box.

The most important point, however, is that in spite of the occurrence of faults, the switching facilities have always allowed a re-arrangement of the equipment to maintain continuity of operation. In fact, radar information has always been available, on at least two out of three displays. This is the justification of the policy of duplication and the switching of units.

In the development of such equipment with switching facilities a number of problems are met, not all of them electronic. Part 2 will discuss some of these problems—and how the equipment has been developed to overcome them.

PART 2

Practical switching arrangement for integrated marine radar system with particular reference to an installation on board S.S. *Canberra*

7. Introduction

The radar equipment installed on S.S. *Canberra* consists of two 10-ft slotted waveguide aerials, two transmitters, two sets of generators, two true plot displays, one of which is coupled to a bright display photographic projector, and a 12-in. relative motion display (Fig. 5).

The inter-switching of these units is carried out in three switch-boxes known as:

(i) transmitter and aerial selector

(ii) generator selector

(iii) slave selector

The slave selector also incorporates an indicator panel which shows the display configuration in use at the time.

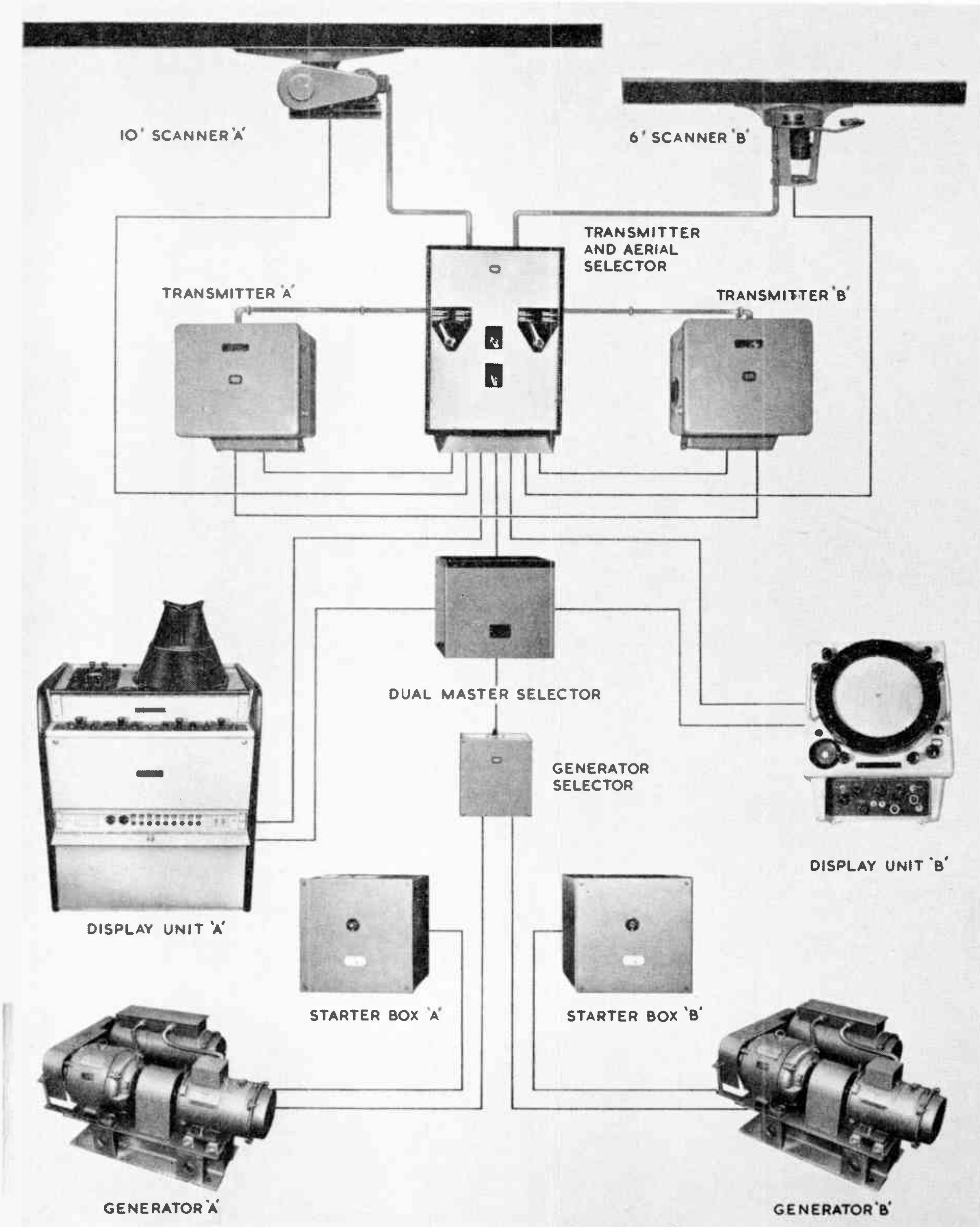


Fig. 5. Two complete and independent radars with inter-switching. Radar 'A' includes 10-ft scanner and 16-in. true-motion display; radar 'B' has 6-ft scanner and 12-in. relative-motion display.

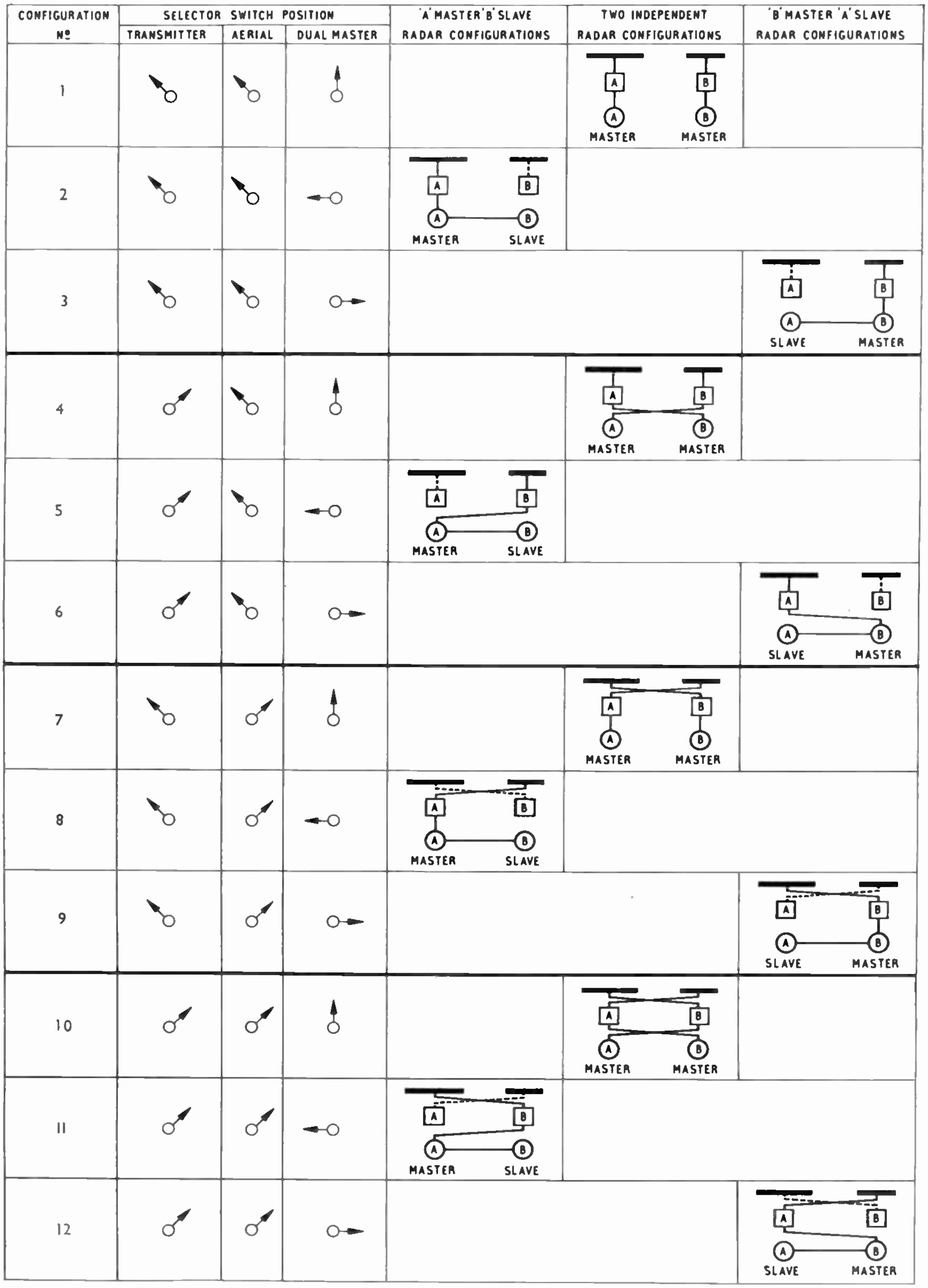


Fig. 6. Main operational configurations. The generators and generator selector switch have been omitted for clarity.

8. Selection of a Display Configuration

There are twelve main operational configurations available, using the two aerials, two transmitters and both true plot displays (Fig. 6). For convenience, the aerials are known as aerial 1 and aerial 2, and the transmitters as transmitter 1 and transmitter 2.

The method of selecting any desired configuration is to determine which aerial is to be used with which transmitter and then which display will operate with the combination of aerial and transmitter already selected. At this stage of the selection, selecting, for instance, aerial 1 with transmitter 1 automatically selects aerial 2 with transmitter 2. Likewise if the aerial 1 transmitter 1 combination is selected to operate with the true plot display then the aerial 2 transmitter 2 combination will be automatically selected to operate with the bright display. If, however, it is desired to operate both true plot displays from the same aerial and transmitter, then this is achieved by deciding which of the displays is to be master, and then selecting this by means of the master selector switch. Throughout the current Type 14 range of equipments, control of the clutter, l.o. tuning and pulse length is carried out in the transmitter. The master display, therefore, controls these functions, all other operational controls on the displays being completely independent of the master display. With this single transmitter and aerial combination selected, it is possible by means of the test selector switch to run the transmitter and aerial not in use for test purposes should a fault develop on either of these units.

9. The Transmitter and Aerial Selector

The initial conception of the transmitter and aerial selector was to have one control knob for each selector control. This entailed a mechanical linkage between the individual switches in each bank, plus a system of interlocks to ensure that only usable combinations of controls were able to be selected. Due to the widely varying currents to be switched, plus the number of poles—one switch had 77 poles involving approximately 300 contacts—different types of switch had to be used.

9.1. *Switch Inter-linkages*

Wafer switches were used for low current switching, and for the heavy current switching, heavy duty 'cooker' type switches were used. Previous experience of coupling wafer switches had shown that to obtain reliable operation over a long period of time, the linkages must have a certain amount of backlash in them, so that in an operating condition each individual switch is indexed by its own mechanism.

In this application, it was necessary not only to couple together wafer switches, but also to couple these to the heavy duty switches and to the waveguide switch. The indexing of the heavy duty switches was 60 and 90 deg, the waveguide switch 90 deg, and the wafer switches 30 deg. These differences in indexing could easily be allowed for by simple reduction gearing into which could also be engineered a certain amount of the desired backlash. Difficulties with this type of coupling arose due to the difference between the number of degrees rotation required at the control shaft to operate the switch, and the final indexed position.

The operating mechanism of a quick make-and-break switch consisted of a contact assembly on its shaft which was indexed by the action of two spring-loaded rollers bearing on a star-shaped cam. To this shaft was connected one end of a spring, the other end of which was connected to the control shaft. Rotation of the control shaft caused this spring to be wound up until the torque produced overcame both the torque and friction required to operate the indexing mechanism. At this point, the indexing mechanism and its contact assembly rotated rapidly to their new position, unwinding the spring. Both the friction of the contact assembly and the rate of rotation of the control shaft affected the 'operate' condition of the switch. A quick make-and-break heavy duty switch with 90 deg indexing can require up to 135 deg rotation of the control shaft to operate it. Even then, it did not necessarily index at 90 deg exactly. At the other extreme, the waveguide switch needed accurate shaft rotation to within 2 deg to ensure precise alignment of the waveguide sections.

Whilst it was possible to produce a system of couplings and interlocks to satisfy these requirements, it was felt that it would be unnecessarily complicated mechanically, very costly to produce and, bearing in mind that this unit was primarily designed to increase the reliability of the installation, it may well have helped to defeat its aim.

9.2. *Final Layout*

An alternative scheme was produced whereby instead of having just the one control knob for each selector, as originally envisaged, each switch has its own control knob, with gearing where necessary such that in any horizontal row of switches the number of degrees rotation for each control knob is the same. This means that for any given operational configuration, the control knobs in any one horizontal row are parallel. Figure 7 shows the control panel with the controls aligned in this manner.

The disadvantage of this switching arrangement was that the configurations could not be changed whilst the equipment was in operation. To prevent

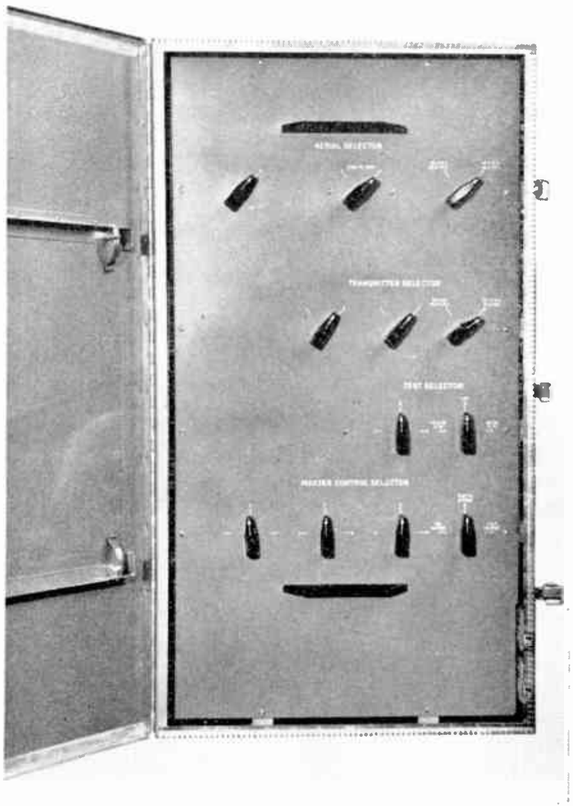


Fig. 7. Exterior view of transmitter and aerial selector showing alignment of controls.

this happening, electrical interlocks are positioned in the unit so that when the door is opened, the equipment is automatically switched off by stopping the generators.

The interlocks can be overridden if it becomes necessary to carry out any servicing of the equipment with the door open. Other than for the waveguide switch, the only damage that can result from trying to change a configuration with the interlocks overridden, is to blow a fuse. Should the waveguide switch be operated with the modulators firing, this could well result in permanent damage to the magnetrons. To prevent this, an additional interlock is installed on the waveguide switch. This switches off the modulators before the switch is rotated and does not enable them to be switched on again until the switch has completed its travel.

9.3. *Lock and Signal Switching*

Lock and signal switching is carried out using aerial switching relays mounted on milled aluminium blocks. The internally-milled channels provide adequately screened interconnection of the relays. This gives reasonable crosstalk figures of 40 dB at the intermediate frequency of 60 Mc/s with an insertion loss of less than 1 dB. When a signal supply is not in use, it is terminated in an impedance equivalent to the input impedance of the receiver, so preventing mismatching of the line affecting other receivers using the same source.

Figure 8 shows the internal arrangements of the transmitter and aerial selector. The lock and signal switching units can be seen at the top right-hand corner. The waveguide switch and its associated interlocks is in the centre of the top row. The interlocks which switch off the generators when the door is opened, can be seen at the bottom of the right-hand side.

10. Slave Selector and Indicator Panel

A chart is provided in the door of the transmitter and aerial selector unit, showing the positions of the switches for any of the twelve operational and eight additional test configurations. The configuration in use is also displayed on the display console by a system of coloured lights on an indicator panel. Also incorporated in this indicator panel is the slave selector. This houses a three-position switch which determines whether the 12-in. display shall be a slave to the true plot display, the bright display or off altogether. If the 12-in. display is inadvertently selected to operate with a display which itself is

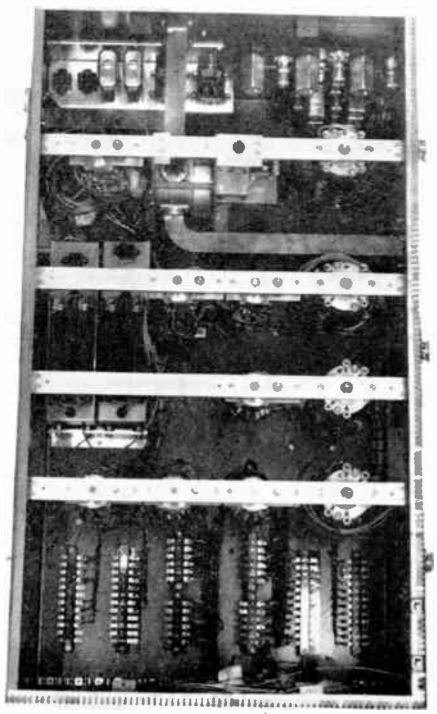


Fig. 8. Interior layout of transmitter and aerial selector.

slave to the other true plot display—this is in a single transmitter and aerial configuration—then it will automatically operate as slave to the master display.

11. Generators

Two sets of twin generators are installed, each set comprising two standard 1 kW generators driven from one 10 h.p. three-phase motor mounted on the same bed. Each generator has its own magnetic amplifier voltage regulator and the driving motor is provided with overload and single phasing protection. Either of these sets can be used to operate any of the configurations previously discussed, the actual set in use being determined by the generator selector unit. As with the transmitter and aerial selector, a test switch is incorporated for running the standby generator for test purposes. If, for any reason, it is desired to change over generators without losing radar information, this can be done by running the standby generator before switching over.

11.1. *Phasing of the Generators and its Effect on Mutual Interference*

In order to minimize the mutual interference between the two transmitters, these are fired at approximately 180 deg out of phase. This is achieved by antiphasing the generators which supply the modulators. Each generator is driven from the motor by its own toothed 'power grip' belt, which ensures that there is no mechanical slip between the generators. The two motor pulleys and one generator pulley are keyed to their shafts, the other generator pulley being held to its shaft by means of a taper grip collet. This enables the relative positions of the generator shafts to be adjusted on test, and then locked in position, thus ensuring that the modulators fire in antiphase. Having carried out this initial setting-up, both the generators and the pulleys are marked, so that in the event of belt failure, no setting up would be required when fitting a new belt. Figure 9 shows a generator layout with the cover removed in which the taper grip pulleys and belts can be seen.

This antiphasing ensures that there is no mutual interference for the first 35 miles range of radar information. At this point, a ring is noticeable on the displays, due to the firing of the other modulator. This is followed by any echoes, about 40 dB down from any targets in close proximity to the ship. Whilst this may seem a disadvantage, in practice it is seldom noticeable. If this type of interference is noticeable, it is easily recognized, due to the fact that the echoes appear to be broken up and 'noisy', and will rotate slowly on the display owing to the aerials not being synchronized. With a single transmitter and aerial configuration there is no interference up to the equipment's maximum range of 48 miles.

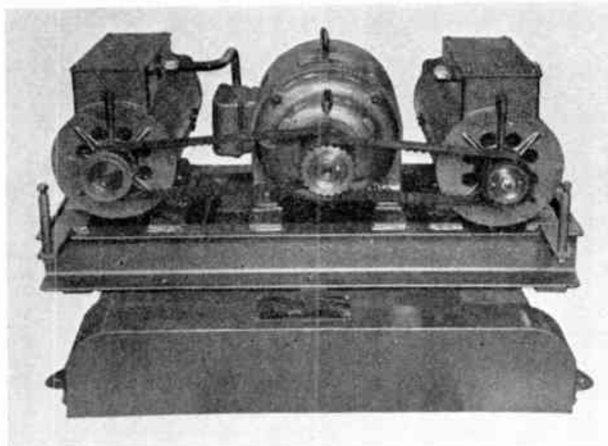


Fig. 9. The belts and pulleys used to ensure correct phasing of the generators.

12. Crystal Protection

Due to the close proximity of the scanners on the mast, it is necessary to afford some protection to the crystal in a transmitter which is not in operation. Normally, under these conditions, the t.r. cell would not be primed. On S.S. *Canberra*, however, the t.r. priming supplies for both transmitters are interconnected so that if either of the equipments are switched on, both t.r. cells are primed. This affords sufficient crystal protection and is shown by the fact that, when a crystal was removed from one of the transmitters after 2200 hours' operation test results show that it had not deteriorated more than one would expect for this operating time in a single installation.

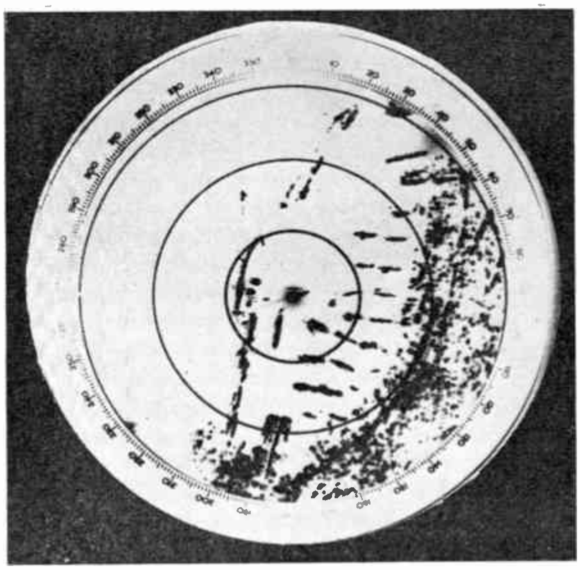
13. Displays

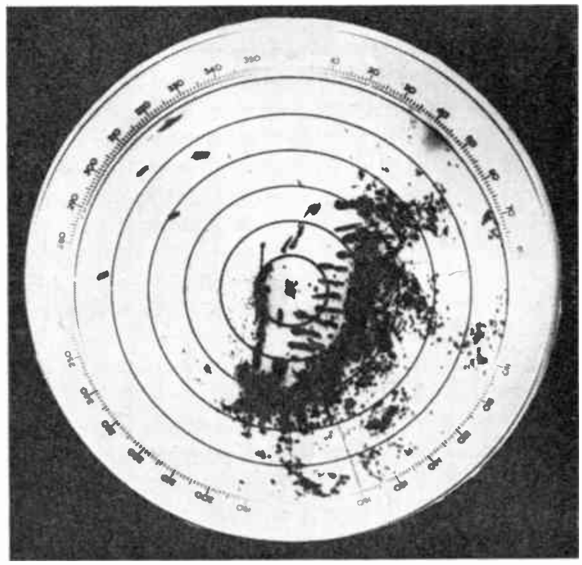
The slave display is a standard 14/12 relative motion equipment with auto-alignment to the ship's head-up presentation. Facilities are also provided for North-up stabilization using a standard compass resolver unit. The head-up/North-up switch is incorporated in the slave power unit alongside the display.

The type 16P which is a 16-in display true plot equipment is standard with a true plot unit, having a speed range up to 40 knots; auto-alignment to either head-up or North-up presentation is included.

The bright display unit is basically the same as the type 16P except for the time-base, which was considerably modified to give the special linearity necessary when used with a photographic projector.† All the electronic units in the two displays are directly interchangeable except for the two time-bases.

†S. R. Parsons, "The application of rapid access photographic techniques to radar display systems", *J. Brit.I.R.E.*, **24**, No. 3, p. 213, September 1962.





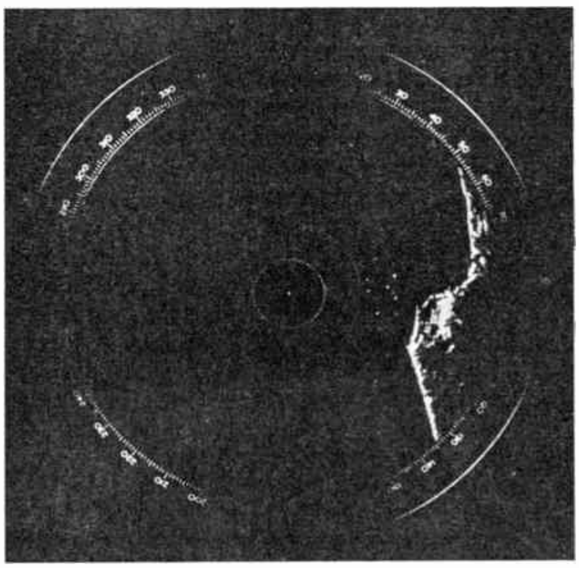


Fig. 10. Radar display of the approach to Colombo.
(*top left*) $\frac{3}{4}$-mile range—negative process.
(*top right*) $1\frac{1}{2}$-miles range—negative process.
(*lower left*) 6-miles range—positive process.

The cathode-ray tube used is a 3-in diameter flat-faced tube with a blue flash, orange after-glow phosphor. The use of a flat-faced tube gives rise to tangential distortion of the trace, causing a non-linear picture to be projected on to the screen. This was overcome by placing a plano-convex lens, with a critical radius of curvature, between the cathode-ray tube and the camera lens. This lens corrected the distortion such that the projected picture was linear to within 1% of full scale deflection. The use of a lens between the tube and camera is to a certain extent a disadvantage due to the fact that there are two more faces to cause internal reflections. Due to the high static field around the face of the tube, there is a tendency to get a film of dust particles and chemical fumes on the lens, causing loss of light, contrast and consequently definition.

In a later installation, instead of a correction lens, a cathode-ray tube was used having an optimum radius of curvature for minimum distortion of the projected picture.

Figure 10 shows prints from the film, processed on board ship, of the radar display of the approach to Colombo. The larger scale pictures give an excellent reproduction of shipping moored head to stern in the harbour.

Figure 10(*c*) shows the positive processing system which can be used if it is preferred to the negative. It was thought that this would be useful for night operation, but this proved to be a mistake. The picture was still too brilliant under dark working conditions, and it has been found that the best solution is the negative picture, but with a deep red filter in front of the projection lens. This agrees with Captain Wepster's† practice of using a red light to illuminate the dials of instruments on a ship's bridge at night, for the same reason—to reduce glare and to prevent the loss of night adaptation of the operator's eyes.

14. Conclusion

By fitting dual display equipments greater reliability often at lower costs, can be obtained from inter-switching the individual units. This leads to the

† A. Wepster, "The arrangement of navigation equipment in modern cargo vessels", *J. Inst. Navigation*, **15**, p. 241, July 1962.

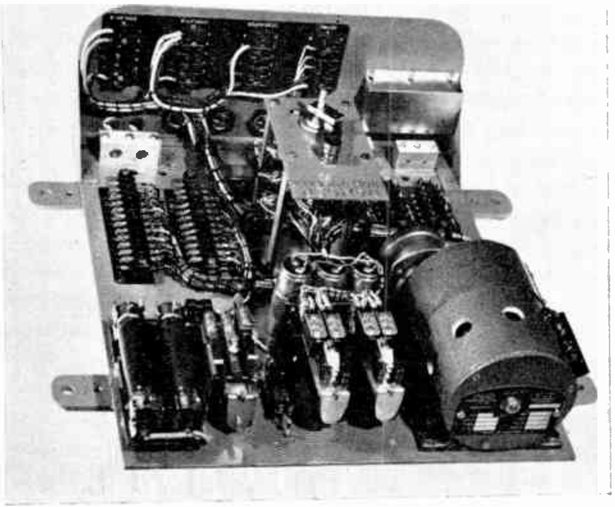


Fig. 11. The internal layout of a switch-box designed for dual display operation, where either display can operate singly or as master of the other display. Switched ship's head-up or North-up presentation is provided on each display.

development of switch-boxes which enable the user to duplicate either true plot or relative motion displays, transmitters, aerials and generators. It is desirable both from the manufacturing and operational aspect that the interswitching of similar sub-units should be carried out in separate boxes. The use of individual switch-boxes makes the switching sequences more comprehensive to the operator and obviates the need for interlocking. Figure 11 shows such a unit.

It would then be possible to have any of these boxes with its associated dual units without using any of the other switch-boxes, i.e. one could have a configuration using one aerial, two transmitters, one display, one generator and one dual transmitter switch-box, or one aerial, one transmitter, two displays (either true plot or relative motion, or one of each), one generator and one dual display switch-box. It is unlikely that one would duplicate such units as the aerial and generator without first duplicating the transmitters and the displays.

The advantage of such a system of switch-boxes is that it enables the user to select a radar system which suits his particular requirements, using the standard units, at the same time keeping the number of sub-units manufactured to a minimum.

15. Acknowledgment

The authors wish to thank the Directors of Kelvin Hughes Division, S. Smith & Sons (England) Ltd., for permission to publish this paper.

Manuscript first received by the Institution on 31st May 1962 and in final form on 6th November 1962 (Paper No. 840/RNA19).

DISCUSSION

Under the chairmanship of Captain F. J. Wylie, R.N.(Retd.)

Mr. L. F. North: What percentage of faults shown against the display unit were in fact due to the receiver, since the latter is incorporated in the transmitter unit in certain other makes of radar equipment?

If complete twin, independent radars were fitted, what percentage increase in availability would be obtained by introducing inter-switching, bearing in mind that the introduction of the inter-switching apparatus could in itself be a source of additional faults?

The Authors (*in reply*): In the Type 14 system the receiver is in two units, the first in the transmitter consisting of 5 stages with a gain of about 40 dB, and being connected by coaxial cable to the second unit in the display. This consists of six stages with a gain of about 60 dB, and also includes a two stage video amplifier. On the 40 ships in the test there were three valve failures in this second unit, i.e. three minor faults out of a total of 61. The balance of figures would not be seriously changed if these faults were transferred to the transmitter.

The second point has not been analysed in detail since the figures indicate that duplication of the aerial and generator will have a negligible effect on system reliability. Duplication is justifiable, however, on the grounds that however unlikely a fault may be, its occurrence at a critical moment could have serious consequences. There are also operational advantages in fitting two scanners. At the date of publication of this paper switch faults still appear negligible.

Mr. P. A. Deegan: What type of ships made up the sample of 48. Were they all, say, ocean going or alternatively coastal craft. Is there a different in fault incidence between, coastal craft and ocean going vessels?

The Authors (*in reply*): Of the 40 ships in the survey, 16 were trawlers, while, of the remaining 24, about one half were coastal and one half deep sea types. The difference in fault rates endorses the view that equipment lasts longer when run continuously than when it is frequently switched off and on again.

Major J. G. Cochran: Could Mr. Harrison indicate what monitoring facilities have been provided in order to provide ease of maintenance by relatively unskilled personnel? It surely follows that reliability must be related to the time taken to cure a fault.

The Authors (*in reply*): The extent of the monitoring facilities is always a problem. We have a small meter indicating power supply, crystal current, and local oscillator tuning. The more important fuses have neon indicators. A performance monitor produces a mark on the p.p.i. whose range indicates that the transmitter-receiver loop gain is normal, or independently that the standard level of power is being radiated, thus checking the wave-guide and scanner.

Using the fault chart illustrated (Fig. 3) and the hand-book, the average navigator can deal with minor faults like valve failures. The man familiar with the equipment and the handbook will do this quickly and thus improve the operational reliability of the equipment.

Mr. J. H. Beattie: Do you use preventive maintenance to improve reliability?

The Authors (*in reply*): Our standard contract service system entails visits to ships by service personnel, and routine tests of equipment whether reported faulty or not. Comparison with standards may indicate the desirability of renewal of a valve, crystal, or magnetron, etc. Regular routines are established for commutator inspection and cleaning, bearing greasing, etc. Our figures indicate that equipments in ships on contract maintenance are at least twice as reliable as in those calling for service when a fault appears.

Mr. I. Davies: The figures quoted by Mr Harrison show, pro rata, a higher faults aggregate than my experience with similar equipment shows. It would be interesting to know if his analysis includes the service reports from Continental agents as I have found that such reports for each visit detailed a number of faults which were non-existent and were consequently very misleading. Does Mr Harrison accept all the "faults" so reported? Again, Continental service people have a habit of including repair of 'dry joints' on practically all reports, and attention to other details which if read by non-technical personnel could produce a bad impression of a manufacturer's products, which are to my knowledge, well constructed and do not suffer from such defects. I have also been present during guarantee repairs when a single fault was found but the subsequent report would describe more than one.

Does Mr Harrison's analysis cover the use of equipment by foreign crews who cannot read the technical manual in English?

In my experience, I would endorse Mr Harrison's aim as a designer in making a radar set as simple as possible to achieve high reliability. I have encountered a particular radar set which had run for two years without service attention.

The Authors (*in reply*): In order to have complete figures on the test, we included only ships which were serviced by our own service department or certain selected foreign agents. I fully agree that fault reports from some sources require translation of more than the mere language. Although our manuals are produced in languages other than English, I could not be certain that in every case the crew had a manual in their own language.

Mr. W. E. Taylor: Is the information from fault analysis used to prepare maintenance schedules?

The Authors (*in reply*): Information from fault analysis is used in three ways:—

(a) To modify maintenance schedules already in use.

(b) To initiate modifications to current production equipment to reduce fault liability. The effectiveness of this use is proved by the fault rates on certain units in S.S. *Canberra* (which were of a later production date than those of the original test) being lower than predicted.

(c) In the design of new equipment. Although the fault rate exhibits random fluctuations the general trend is a progressive reduction from the original type-approved set to one in current production.

Mr. T. G. Clark (*Communicated*): What confidence level is assigned to the probability of failure figures given and can the figures be extrapolated to a 1000 hour period in order to afford direct comparison with normal statistical data?

The Authors (*in reply*): It is not possible to give a general answer to the first question. For the Poisson distribution the mean value z is equal to the variance so the standard deviation is equal to the square root of z, from which the confidence level for any particular case may be evaluated.

The figures may be used to compute probability of failure for any period by calculating the mean value z for that period. Thus to revise Section 4.2 for 1000 hours we write

$$\text{Average number of faults in 1000 hours} = z = 0{\cdot}431 \times \frac{1000}{300} = 1{\cdot}44$$

$e^{-z} \simeq 0{\cdot}24$

Probability of success $= e^{-z} = 0{\cdot}24$

Probability of 1 fault $= ze^{-z} = 0{\cdot}35$

Probability of 2 faults $= \dfrac{z^2}{2!} \times e^{-z} = 0{\cdot}25$

etc.

Thus $S = 0{\cdot}24$; $F = 0{\cdot}76$.

U.D.C. 621.373

The Performance of an Ammonia Maser with Two Resonators in Cascade

By

D. C. LAINÉ, Ph.D.†

AND

R. C. SRIVASTAVA, Ph.D.‡

Presented at the Symposium on "Masers and Lasers" in London on 2nd January, 1963

Summary: An ammonia maser is described in which an additional resonant structure is inserted between the microwave cavity and the electrostatic separator. Observation of the power level of the microwave oscillation is shown to be a sensitive means of detecting various types of molecular transitions in the auxiliary resonator.

It is found that when this auxiliary resonator is a microwave cavity which can be tuned across the molecular resonance, the power level of the oscillation in the maser cavity varies in a more complex way than would be expected on consideration of population differences alone. In particular, a beat signal of frequency 4–5 kc/s superimposed on the normal oscillation signal is observed for a critical detuning of the additional cavity from the molecular resonance frequency. The case when the auxiliary cavity takes the form of a multimode ultramicrowave resonator is also considered. It is suggested that transitions of ammonia molecules amongst the various rotational levels should be observable through changes in the amplitude of microwave oscillation.

1. Introduction

The ammonia maser[1,2] is now well-known as the forerunner of the numerous devices using the principle of stimulated emission from atoms and molecules for amplification and generation of electromagnetic radiation.

One of the main interests in the ammonia maser has been derived from the exceptional frequency stability of the oscillation signal in the region of 24 Gc/s. The various factors that limit the reproducibility of the maser frequency have been studied in considerable detail by many workers and considerable success has been achieved in the control of frequency pulling effects. The ammonia maser may now be considered to compete favourably both in long and short term stability[3] with other types of 'atomic clock'.

Only two other types of beam masers have been made to oscillate since the publication of the details of the operation of the ammonia maser oscillator in 1954.[1] The first of these was the atomic hydrogen maser[4] which was operated in 1960 as an oscillator at 1·420 Gc/s. The other maser used a beam of hydrogen cyanide molecules[5] to produce an oscillation at 88·6 Gc/s. Several other types of beam masers have been operated as spectrometers and it is probable that the wider applications of the beam maser technique lie in this direction.[6]

In the course of experiments using the oscillating ammonia maser as a detector of molecular transitions occurring in an auxiliary resonator, it has been found that the experimental results are more complex than would be suggested by a simple consideration of the relative population of molecules in the two inversion levels corresponding to the $J = 3$, $K = 3$ rotational state. Details of apparatus, technique and principal experimental results are reported in this paper. Finally a proposal is made to extend these experiments to use the oscillating ammonia maser as a spectrometer or a detector of radiation in the ultramicrowave region using absorption or stimulated emission between rotational states of ammonia molecules.

2. The Maser as a Detector of Molecular Transitions in an Auxiliary Resonant Structure

The amplitude of the microwave oscillation in an ammonia maser is strongly dependent on the separator voltage, the molecular flux and the cavity tuning. For a given beam flux and cavity tuning an increase in the separator voltage causes a larger number of molecules in an upper energy state to pass into the microwave cavity and contribute energy to the oscillating field by the process of stimulated emission. The maser oscillation amplitude as a function of separator voltage for a small value of beam flux is shown in

† Physics Department, University of Keele, Staffordshire.

‡ Formerly Physics Department, University of Keele; now at Physics Department, University of Lucknow, India.

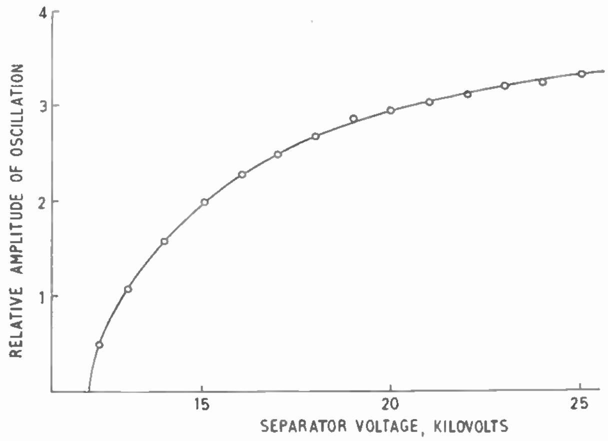


Fig. 1. Relative amplitude of maser oscillation as a function of separator voltage.

Fig. 1. It is seen from this curve that at higher values of separator voltage the output power from the maser tends towards a maximum value which indicates that a much larger proportion of the molecules in the upper energy state are separated out to contribute to the level of oscillation.

If the separator voltage and cavity tuning are kept constant then the variation of the amplitude of oscillation as a function of ammonia gas pressure behind the gas inlet nozzle is of the form shown in Fig. 2. The amplitude of oscillation is limited by saturation effects occurring in the microwave cavity where molecules once having emitted re-absorb energy before they pass out of the cavity. At high beam pressures the collimation of the molecular beam deteriorates due to molecular collisions within the nozzle and separator and also at the entrance to the maser cavity. The net effect of molecular scattering is to increase the background pressure along the beam axis so that the number of active molecules passing into the microwave cavity is reduced and consequently the amplitude of oscillation falls. It is seen from Figs. 1 and 2 that the amplitude of oscillation (which depends on the number of molecules in the upper energy state) is increased with higher values of separator voltage, and also by increasing beam flux up to the limit set by molecular scattering. The amplitude of oscillation is most sensitive to changes in the number of molecules in an upper energy state at a low value of beam flux as is seen in Fig. 2 and under such conditions molecular transitions in an auxiliary resonator placed between the normal oscillating cavity and the state separator can be detected most easily. For example, if the auxiliary resonator takes the form of a tuned microwave cavity in which a radiation field is present at the molecular resonance frequency, then a change in the amplitude of microwave oscillation is expected in the maser cavity owing to the robbing of molecules in the upper energy state to produce stimulated emission in the auxiliary cavity.[7] The inducing field can of course arise solely from thermal radiation within the auxiliary cavity since any radiation with frequency components within the molecular line width is amplified by maser action and this can lead to oscillation if the total power emitted exceeds that absorbed by the system. If self-sustained oscillations are maintained in an auxiliary cavity then a very large decrease in the amplitude of oscillation occurs in the normal maser oscillator cavity as will be shown later when the experimental results are considered.

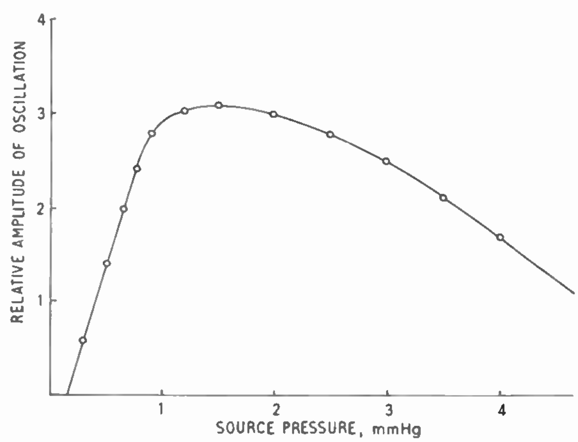


Fig. 2. Relative amplitude of maser oscillation as a function of ammonia source pressure.

If the auxiliary cavity takes the form of a multimode ultramicrowave resonator, then molecular transitions induced between levels corresponding to different rotational states of the ammonia molecule should be observable via the microwave oscillation if a sufficiently large ultramicrowave radiation field is present. Such a field could arise from thermal radiation within the multimode resonator or be produced from an external source such as a mercury lamp. A discussion of the use of the ammonia maser oscillator as a detector of stimulated emission in the ultramicrowave region is given in Section 5.

3. Apparatus

In an ordinary ammonia maser, the distance between the state separator and the microwave cavity is usually between 1 and 2 cm. If this distance is increased in order to insert an auxiliary resonator structure then the divergence of the beam will cause a reduction in the number of molecules entering the cavity provided that the beam flux is kept constant. This distance has been increased to 8 cm in the present apparatus. In order to obtain strong oscillations in the maser cavity good collimation of the beam is necessary and this is achieved by using a honeycomb nozzle 5 mm long and 4 mm diameter made from klystron grid bar stock. The state separator is constructed with eight parallel wires arranged around the circumference of a circle 1 cm in diameter. A voltage between 15 and 25 kV is applied between adjacent wires for normal operation of the maser. Ceramic is used in preference to other materials to minimize breakdown in the insulation of the separator assembly. A 4-in diffusion pump is used to maintain the background pressure to below 5×10^{-6} mm of mercury.

The separator voltage may be plotted as a function of the nozzle pressure necessary to produce threshold oscillations for several values of the distance between cavity and separator and the results are shown graphically in Fig. 3. The curves A, B, C correspond to different distances between the cavity and a 12 cm long separator but without the auxiliary resonator in position. Curve D is obtained with a distance of 8 cm between cavity and separator, but in this case a separator 8 cm long is used and the beam is passed through an auxiliary microwave resonator which is not tuned to the molecular resonance. The minimum separator voltage for threshold oscillations is seen to be reduced in curve D. This is attributed to improved efficiency of the shorter separator at low gas pressure due to a larger proportion of the molecules effusing from the nozzle reaching the maser oscillator cavity before the effects of molecular scattering in the auxiliary resonator become appreciable. It is found that satisfactory operation of the maser can be obtained with a distance between cavity and separator up to 10 cm. This information has been incorporated in the design of the dual resonator maser shown in section in Fig. 4. In this diagram the auxiliary resonator is shown as a microwave cavity, but the apparatus is designed so that larger multimode resonator structures can also be used. The cavities used in the experiments described in this paper are thermally tuned by means of a bifilar wound heater. Copper wire is used as the heater element so that its high value of temperature coefficient of resistance can be used for the electronic stabilization of cavity temperature.

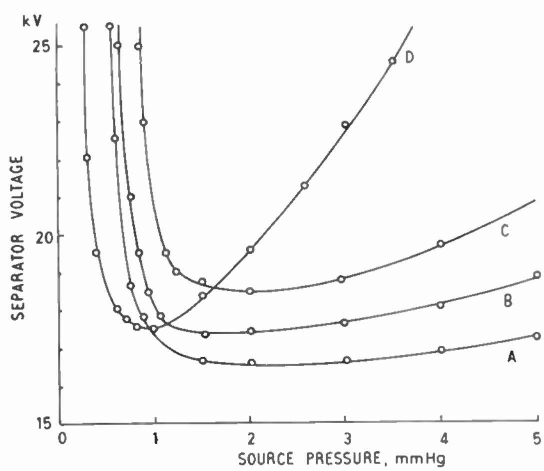


Fig. 3. Threshold voltage for maser oscillation as a function of source pressure.

Distance between cavity and separator, A 1 cm; B 3·5 cm; C 6 cm; D 8 cm.

Separator length, 12 cm for curves A, B and C; 8 cm for D.

Curve D was obtained with an auxiliary microwave cavity in place.

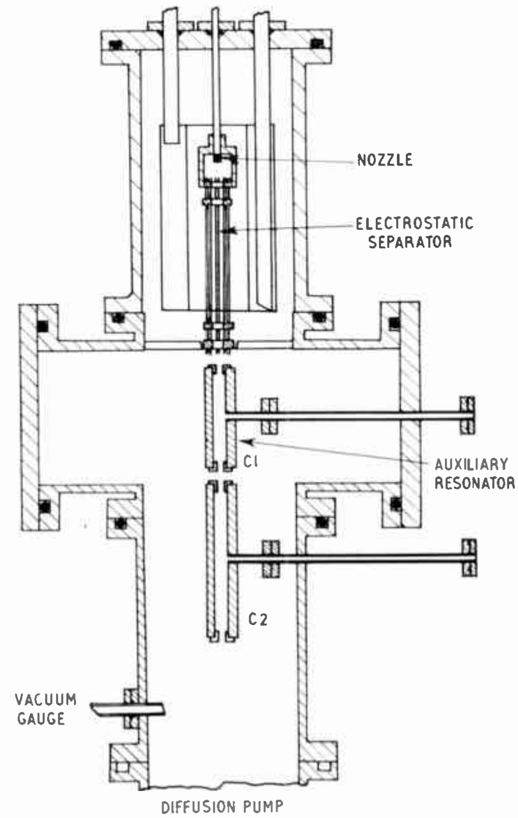


Fig. 4. Section of ammonia maser with two microwave cavities in cascade. The e.h.t. lead-in is not shown.

The amplitude of the oscillation in the maser cavity is monitored by a simple superheterodyne system as follows. The power output of the local oscillator K-band klystron is frequency modulated over several megacycles by a 50-c/s sawtooth voltage applied to the reflector and the resultant beat signal (produced by non-linear mixing of the maser and local oscillator signal in a 1N26 microwave detector crystal) is also frequency modulated. The klystron frequency is adjusted so that this beat signal sweeps through the 2 Mc/s passband of an i.f. amplifier tuned to 30 Mc/s and the demodulated output is used to display the i.f. amplifier frequency response on an oscilloscope as shown in Fig. 5. The amplitude of the beat signal

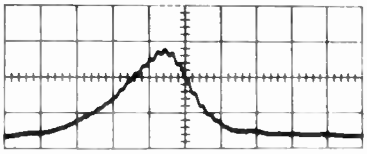


Fig. 5. Oscilloscope trace of the i.f. amplifier bandpass where the amplitude is proportional to the level of maser oscillation. Amplitude modulation is shown superimposed on the trace.

is proportional to the amplitude of oscillation in the maser cavity. A portion of the signal fed into the oscilloscope is passed into a low gain a.c. amplifier and the amplified 50-c/s signal is passed into a full-wave bridge rectifier using four silicon diodes. The d.c. output from this bridge is taken to a 0–5 mA linear pen recorder so that the pen deflection is then proportional to the amplitude of the maser oscillation signal. The microwave system and auxiliary apparatus are shown in Fig. 6.

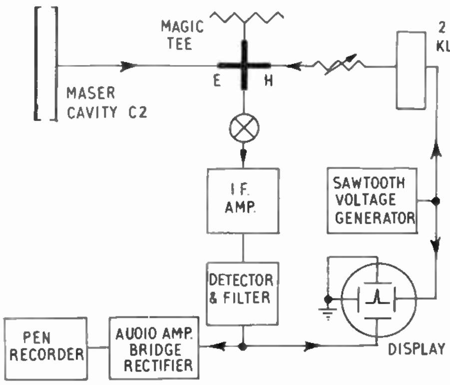


Fig. 6. Block diagram of apparatus used to monitor the amplitude of maser oscillation.

4. Microwave Cavity as the Auxiliary Resonator

The results obtained using a microwave cavity as an auxiliary resonator, henceforth abbreviated to C1 ($Q \simeq 4000$, E_{010} mode, length 7 cm), placed between the oscillating microwave cavity C2 ($Q \simeq 5000$, E_{010} mode, length 10 cm) and the separator may be seen by reference to Fig. 7 where a pen recording of the level of microwave oscillation in the cavity C2 is shown as a function of the auxiliary cavity tuning. These recordings are obtained when C2 is tuned to the molecular resonance frequency. Since thermal tuning is used, the resonant frequency of C1 can be made to pass slowly through the region of interest by passing current through the cavity heater. The resonant frequency of the cavity then becomes a function of time, so that a pen recording of the amplitude of oscillation becomes a function of frequency. No special attempt has been made to obtain a linear time-cavity frequency relationship. Recordings for both heating and cooling are shown in Fig. 7.

As the cavity C1 is tuned nearer to the molecular resonance, an increasing number of molecules will be stimulated to radiate energy in this cavity. Cavity C2 is therefore gradually robbed of molecules which would otherwise contribute to the amplitude of oscillation. If the oscillation level in C2 is simply dependent upon the net excess number of molecules in the upper energy state reaching it from C1, then it would be expected that the oscillation in C2 would be gradually reduced from the normal oscillation level to a minimum value and back again as C2 is tuned through the region of the molecular resonance frequency. However, it is seen from Fig. 7 that the amplitude of oscillation shows not just one but three minimum values and also two maximum values within the tuning range AD where oscillation occurs in C1.

A further interesting phenomenon occurs in the regions AB and CD of Fig. 7, where the amplitude of oscillation changes rather abruptly. This effect, which was first observed by Higa,[8] takes the form of a low-level amplitude modulation which is shown on the oscilloscope trace in Fig. 5. The frequency of this modulation is found to be in the region of 4 to 5 kc/s; the precise value depends on the quality factor of C1 and the resonant frequencies of C1 and C2. The modulation frequency can also be changed by a few hundred cycles by alteration of either the beam pressure or the separator voltage. The amplitude modulation corresponds to the difference frequency between oscillation in C2 (usually tuned to the centre frequency ν_M of the molecules) and the frequency pulled oscillation in C1 tuned away from the molecular line. The frequency pulling in a molecular oscillator is given for cavity C1 by the equation[9]

$$\nu_1 - \nu_M = \frac{Q_{C1}}{Q_M} \cdot (\nu_{C1} - \nu_M) f(n) \qquad \ldots\ldots(1)$$

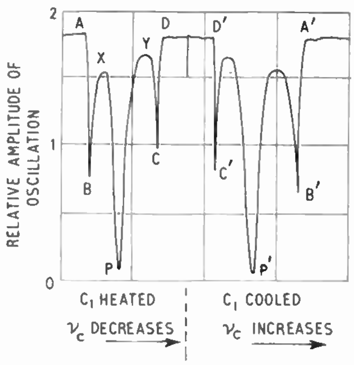


Fig. 7. Pen recording of the relative amplitude of maser oscillation in C2 as a function of frequency of C1 which is tuned thermally (*a*) when heated, (*b*) when cooled. The frequency scale is non-linear. The frequency deviation of C1 between points AD, D′A′ is about 7 Mc/s. Regions where amplitude modulation occurs: AB, CD, D′C′, B′A′.

where ν_1 is the frequency of the oscillation in C1, ν_{C1} the resonant frequency of C1, f(n) is a complex function of beam flux and amplitude of oscillation, Q_{C1} is the loaded quality factor of C1 and finally the Q of the molecular resonance is defined by

$$Q_M = \frac{\nu_M}{2\Delta\nu_M} = \frac{\nu_M}{0{\cdot}89} \cdot \frac{L}{v} \qquad \text{......(2)}$$

where $2\Delta\nu_M$ is the spectral width of the molecular resonance, L is the length of the microwave cavity, v is the velocity of the molecules. For small oscillations in C1, saturation effects can be neglected and the following values can be substituted in eqns. (1) and (2).

$$f(n) \simeq 1, \qquad Q_{C1} = 4000,$$

$$L = 7 \text{ cm}, \qquad v = 6{\cdot}0 \times 10^4 \text{ cm/s}$$

$$\nu_M = 2{\cdot}4 \times 10^{10} \text{ c/s}$$

so that $Q_M = 3 \times 10^6$ and

$$\nu_1 - \nu_M = 1{\cdot}3 \times 10^{-3}.(\nu_{C1} - \nu_M) \qquad \text{......(3)}$$

Thus if the oscillation frequency in C1 is pulled by say 4·5 kc/s, then the corresponding detuning of the cavity C1 from the molecular resonance is about 3·5 Mc/s. The change in the frequency of C1, between points A and D of Fig. 7 should then be twice this value, that is, 7 Mc/s. The measured change in the resonant frequency of the cavity over this range is in good agreement with this value.

Now for cavity C2 to oscillate simultaneously at two frequencies differing by the beat frequency, a means of coupling the oscillation in C1 to C2 must exist. The two microwave cavities are provided with close fitting end caps as shown in Fig. 4 to prevent radiation of microwave power from the ends of the cavities, and in addition, the separation of 1 cm between cavities ensures that any microwave coupling is sufficiently small as to be neglected. However, another means of coupling between the cavities exists via the molecular beam which has been termed 'molecular ringing'.

The molecular beam, having a resonance of bandwidth $2\Delta\nu_M$ has a memory or ringing time of about $\tau = (2\pi\Delta\nu_M)^{-1}$ seconds by analogy with a tuned circuit. Since the beam passes first through C1 where some molecules have already emitted energy to maintain the oscillation in C1, then due to ringing, the molecules continue to emit in C2, at the frequency of oscillation[10,11] determined by C1. The amplitude of the oscillation signal monitored from C2 shown in Fig. 7 as a function of tuning of C1 may now be understood in a crude manner as follows.

First consider that the cavity C2 is oscillating at the molecular resonance frequency ν_M, and that the level of oscillation is constant until the tuning of C1 reaches the value given by point A from the left-hand side of Fig. 7. At this point stimulated emission in C1 is sufficient for an oscillation to build up at the frequency ν_1 as given by eqn. (1). Due to molecular ringing, a fraction of this oscillation signal appears in C2 which mixes with the oscillation signal in C2 at frequency ν_M through non-linear effects in the maser, and an amplitude modulation appears superimposed on the output power as shown in Fig. 5. As C1 is tuned nearer to ν_M, and the net number of molecules left in the upper energy state reaching C2 diminishes, the amplitude of oscillation in C2 falls. The rapid drop in power level is also enhanced by amplification of the ringing signal in C2 which robs the oscillation of frequency ν_M of yet more molecules. At point B, the amplified ringing signal is sufficiently large to quench oscillations in C2 at the frequency ν_M, and it thus becomes the driving field for stimulated emission in C2. The frequency of this field is determined by eqn. (1), so that the frequency of the output power from C2 now follows the oscillation frequency of C1. As the oscillation signal continues to grow in C1, the amplified ringing signal which now provides the output signal from C2, will also grow. However, when the oscillation in C1 starts to saturate (which arises when the number of molecules in upper and lower states are nearly equalized), few molecules are available for amplification of the ringing signal in C2 and the power output falls to a minimum value at the frequency of the molecular resonance denoted by the point P. Further tuning of C1 beyond P repeats the pattern in reverse order.

A further interesting feature of Fig. 7 is the asymmetry of the curve. This has been found to be due to the maser oscillator cavity C2 being slightly off-tune from the molecular resonance frequency. Accurate tuning of cavity C2 for symmetry of this curve could possibly be used as an approximate method for setting the cavity tuning of a maser to the centre of the molecular resonance without reference to a second standard of frequency.

The overall range of tuning of C1 which corresponds to the separation between A and D of Fig. 7 is found to be extended with increased values of separator voltage and beam source pressure. This occurs because of the greater tuning range over which oscillations can be obtained in C1 when a greater number of molecules are available in the upper energy state.

It is found that when the frequency of C1 is set at a value corresponding to the region AB, or CD of Fig. 7 (maser operated at 21 kV, source pressure 1·1 mm of mercury), then the change in the relative amplitude of oscillation in C2 as a function of separator voltage appears as shown in Fig. 8. It is seen that the form of the curve is different from that obtained for normal maser operation shown in Fig. 1. The shape of this curve may be understood qualitatively as a combination of two effects. The first of these is the normal increase in oscillation amplitude in C2, with separator voltage as shown in Fig. 1 until C1 starts to oscillate. At this point the second effect becomes predominant and stimulated emission in C1 reduces the amplitude of oscillation in C2, and at the same time amplitude modulation appears. As the separator voltage is increased, the pattern shown in Fig. 7 expands and the wings AB, CD of the curve move further apart. Since the frequency of C1 is fixed, expansion of the curve can be seen to produce the change in amplitude of oscillations in C2 as the portions ABX, or YCD move outwards with increasing separator voltage.

The theory of the cascaded cavity maser has been considered by Javan and Wang,[7] Serber and Townes[10] and also Wells.[11] A rigorous quantum-mechanical treatment of the problem yields unwieldy solutions to the problem in all cases except when the cavities C1 and C2 are tuned to the molecular resonance frequency. Less cumbersome solutions are being sought which will lead to a more detailed quantitative description of the effects reported in this paper.†

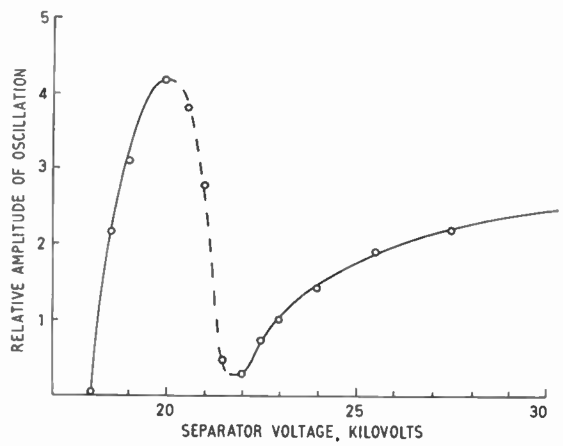


Fig. 8. Relative amplitude of oscillation in C2 as a function of separator voltage. The broken line indicates the region where amplitude modulation occurs.

† *Note added in proof:* Subsequent to the completion of this paper for publication, a detailed theoretical paper entitled "Emission from molecules in a mixed energy state" has been published by N. G. Basov and A. N. Oraevskiĭ (*J. Exptl Theor. Phys.* (*U.S.S.R.*), **42**, p. 1529, 1962; *Soviet Physics JETP*, **15**, p. 1062, 1962). References are made in this paper to other publications on the same topic of which the authors were unaware. Of these, a paper by G. M. Strakhovskiĭ and V. M. Tatarenkov entitled "Radiation of molecules under resonant conditions" (*J. Exptl Theor. Phys.* (*U.S.S.R.*), **42**, p. 907, 1962; *Soviet Physics JETP*, **15**, p. 625, 1962) is of particular interest since experimental results are reported which are equivalent to those shown in Fig. 7 of this paper.

5. Ultramicrowave Cavity as the Auxiliary Resonator

It has already been pointed out that the amplitude of oscillations in an ammonia maser can be influenced by transitions amongst rotational states of the ammonia molecules. The energy levels corresponding to the two lowest rotational states for $K = 3$, are shown in Fig. 9(*a*). The possible transitions between these levels are given by the selection rules

$$\Delta J = 0 \pm 1, \qquad \Delta K = 0, \qquad (+) \leftrightarrow (-)$$

where the upper and lower inversion levels are designated $(+)$ and $(-)$ respectively. The inversion transition designated $(+) \rightarrow (-)$ for the lowest rotational state $J = 3$, $K = 3$ is the usual one for the operation of an ammonia maser. The next pair of inversion levels in the higher rotational state $J = 4$, $K = 3$, has a slightly lower molecular population than the ground state. Since these levels are the most important in the context of this discussion the higher rotational states will be ignored. The inversion frequency of the lower pair of levels is 23·870 Gc/s whilst for the upper pair it is 22·688 Gc/s. The ultramicrowave transitions between levels 1 and 4, 2 and 3 are also permitted and correspond to transitions at approximately 2405 Gc/s and 2358 Gc/s respectively.[12] When the molecular beam passes through the electrostatic separator, the molecules in the lower inversion level of each rotational state are removed as indicated

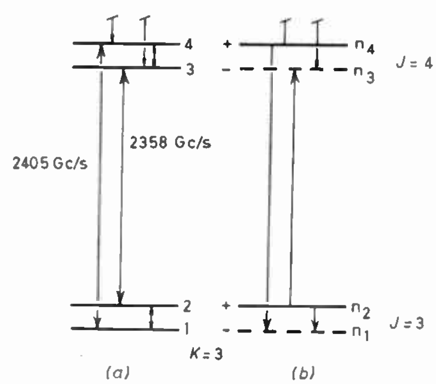


Fig. 9. Rotation inversion energy levels and permitted transitions between the $J = 4$, $K = 3$; $J = 3$, $K = 3$ states of ammonia molecules. Empty levels are shown by broken lines. (*a*) Before separation. (*b*) After separation.

by the broken lines in Fig. 9(*b*). If the beam is then passed through an ultramicrowave resonator either stimulated emission or absorption processes will tend to change the number of molecules in the various energy levels. The amplitude of oscillation in the normal microwave ammonia maser is dependent on the excess number (n_2) of the molecules in the upper inversion state over the number (n_1) in the lower state which pass into the cavity per second, and the amplitude of oscillation is a function of the net beam flux $(n_2 - n_1)$.

If the molecules are first passed into an ultramicrowave resonator tuned to either the transition corresponding to emission or absorption then any background thermal radiation from the cavity itself, or radiation at the appropriate wavelength introduced into the resonator from an external source will cause stimulated emission by inducing molecules to move from level 4 to level 1, thus increasing n_1, or be absorbed by inducing transitions between levels 2 and 3 thus decreasing n_2. In both instances the net beam flux $(n_2 - n_1)$ molecules per second is decreased so that the amplitude of maser oscillation in C2 is reduced. An idea of the sensitivity of the method can be obtained from the following considerations.

If the beam maser spectrometer is operated under optimum conditions of cavity coupling and microwave power level, the minimum change in beam flux $n_{\min}$ molecules per second which can be detected at the microwave frequency ν_0 is given by the expression[6]

$$n_{\min} = \frac{1 \cdot 52 v}{\bar{\mu} L}\left[\frac{VFkT\Delta f}{\nu_0 Q_0}\right]^{\frac{1}{2}} \qquad \ldots\ldots(4)$$

where v is the velocity of the molecules, $\bar{\mu}$ is the average dipole matrix element for the component along the direction of the electric field, k is Boltzmann's constant, F the overall noise figure of the detector and amplifier, Δf the effective bandwidth of the detection system; T, Q_0, A and V are the temperature, unloaded quality factor, cross sectional area and volume of the microwave cavity respectively.

Substituting the following values: $v = 6 \times 10^4$ cm/s, $\bar{\mu} = 10^{-18}$ c.g.s. units, $k = 1 \cdot 4 \times 10^{16}$ erg deg^{-1}, $F = 10$, $\Delta f = 0 \cdot 1$ c/s, $T = 290°$ K, $Q_0 = 2Q = 10^4$, $A = 1$ cm^2, $V = 10$ cm^3, $\nu_0 = 2 \cdot 4 \times 10^{10}$ c/s, then

$$n_{\min} = 3 \cdot 8 \times 10^8 \text{ molecules/second} \qquad \ldots\ldots(5)$$

This sensitivity can be achieved provided that the parameter θ defined by

$$\theta = \frac{\pi E \bar{\mu} L}{h\nu} \qquad \ldots\ldots(6)$$

takes the optimum value of 1·16. In this equation E is the electric field in the microwave cavity and h is Planck's constant. This value of θ gives the condition for the maximum emission of power from $n_{\min}$ molecules.

The electric field which stimulates emission within the maser cavity is usually generated by an external klystron. However, when the maser oscillates, the microwave field is self-maintained and any change in the number of molecules either in the upper or lower inversion states will alter the amplitude of the maser oscillation. The minimum change in beam flux which can be detected by monitoring the level of oscillation by means of a conventional superheterodyne detection system may then be considered to be roughly equivalent to that determined by eqns. (4) and (5). The optimum value of $\theta = 1 \cdot 16$ is obtained in C2 when the ratio of the actual beam flux to the beam flux required to produce threshold oscillations[9] is about 3. If it is assumed that the beam flux is proportional to the pressure of ammonia gas behind the nozzle, then the condition for maximum sensitivity is obtained for an ammonia gas source pressure of about 0·5 mm of mercury, which corresponds to the middle part of the steep portion of the curve shown in Fig. 2.

Now if the minimum detectable change in beam flux given by eqn. (5) results either from emission or absorption between rotational levels of ammonia molecules, then the minimum power absorbed (or emitted) by the beam which can be detected at the ultramicrowave frequency ν_{umw} by observation of the amplitude of the microwave oscillation at frequency ν_M is given by

$$P_{\text{umw(min)}} = h \, . \, \nu_{\text{umw}} n_{\min} \qquad \ldots\ldots(7)$$

substituting the values given previously into (7), then

$$P_{\text{umw(min)}} = 6 \times 10^{-13} \text{ watts} \qquad \ldots\ldots(8)$$

The sensitivity may be even better than this in practice, since no attention has been given to amplification processes in the maser at either the microwave or ultramicrowave frequencies. The use of the maser oscillator in observations on the hyperfine structure of ammonia has been found by Shimoda and Wang[13] to give an improvement on the usual maser spectrometer sensitivity by a factor of about 50. Thus it is probable that the minimum detectable ultramicrowave power given by eqn. (8) may be as low as 10^{-13} or even 10^{-14} watts.

6. Conclusions

It will be seen from the previous sections that the operation of an ammonia maser with two cavities in cascade where molecular transitions occurring in an auxiliary resonator are detected by monitoring the amplitude of oscillations in a normal maser oscillator cavity is a useful variation on beam maser technique. In the case where the auxiliary cavity is a microwave resonator, several complicated effects have been observed and explained on a qualitative basis. It is suggested that the symmetry of some of the experimental curves could be used where the maser is employed as a frequency standard as an alternative means of setting the maser cavity to the centre of the molecular resonance without further reference to other standards. It is also proposed that an extension of the two resonator system to attempt detection of molecular transitions between rotational levels of ammonia should also be possible. It can be seen that the use of such a double resonance method largely transfers the problem of the detection of stimulated emission in the ultramicrowave region of the spectrum (far infra-red) to the microwave region, where the sensitivity of a microwave maser followed by a superheterodyne detection system is available. Detection of stimulated emission by such a technique could then lead to an appraisal of the possibilities of operating a maser as an ultramicrowave oscillator at a wavelength of 0·125 mm as suggested by Barnes.[14]

7. Acknowledgments

The authors wish to express their appreciation to Mr. G. Phillips, Mr. G. S. Waters and Mr. A. M. J. Mitchell of the Signals Research and Development Establishment, Ministry of Aviation, for support and interest in this work. The authors also express thanks to Professor D. J. E. Ingram for the provision of laboratory facilities and for his encouragement. One of us (R.C.S.) is grateful to the Commonwealth Scholarship Commission in the United Kingdom for the award of a Commonwealth Scholarship. Finally, the authors are indebted to the Ministry of Aviation for permission to publish this paper.

8. References

1. J. P. Gordon, H. J. Zeiger and C. H. Townes, "Molecular microwave oscillator and new hyperfine structure in the microwave spectrum of NH_3", *Phys. Rev.*, **95**, p. 282, 1954.
2. J. P. Gordon, H. J. Zeiger and C. H. Townes, "The maser—new type of microwave amplifier, frequency standard and spectrometer", *Phys. Rev.*, **99**, p. 1264, 1955.
3. K. Shimoda and N. Kohno, "Ammonia maser on the 3,2 line as a frequency standard", *J. Appl. Phys. Japan*, **1**, p. 5, 1962.
4. H. M. Goldenburg, D. Kleppner and N. F. Ramsey, "Atomic hydrogen maser", *Phys. Rev. Letters*, **5**, p. 361, 1960.
5. D. Marcuse, "Maser oscillation observed from HCN maser at 88·6 kMc", *Proc. Inst. Radio Engrs*, **49**, p. 1706, 1961.
6. K. Shimoda, "Maser spectroscopy", *R.C. Soc. Ital. Fis.*, **17**, p. 1, 1960.
7. A. Javan and T. C. Wang, "Two cavity maser spectrometer", *Bull. Amer. Phys. Soc.*, **2**, p. 209, 1957.
8. W. H. Higa, "Observations of non-linear maser phenomena", *Rev. Sci. Instrum*, **28**, p. 726, 1957.
9. K. Shimoda, T. C. Wang and C. H. Townes, "Further aspects of the theory of the maser", *Phys. Rev.*, **102**, p. 1308, 1956.
10. R. Serber and C. H. Townes, "Limits on electromagnetic amplification due to complementarity", "Quantum Electronics", Edited by C. H. Townes, p. 233 (Columbia Univ. Press, 1960).
11. W. H. Wells, "Maser oscillator with one beam through two cavities", *J. Appl. Phys.*, **29**, p. 714, 1958.
12. H. M. Mould, W. C. Price and G. R. Wilkinson, "A high resolution study and analysis of the ν_2 NH_3 vibration-rotation band", *Spectrochimica Acta*, **21**, p. 313, 1959.
13. K. Shimoda and T. C. Wang, "New method for the observation of hyperfine structure of NH_3 in a maser oscillator", *Rev. Sci. Instrum*, **26**, p. 1148, 1955.
14. F. S. Barnes, "The feasibility of building beam type masers in the millimeter and submillimeter wave range", "Quantum Electronics", Edited by C. H. Townes, p. 57 (Columbia Univ. Press, 1960).

Manuscript first received by the Institution on 7th November 1962 and in final form on 8th February 1963 (Paper No. 841).

U.D.C. 534.88 : 551.46.018

Profile and Area Echograph for Surveying and Location of Obstacles in Waterways†

By

S. FAHRENTHOLZ, Dr. Ing.‡

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

Summary: The survey ships carry outriggers floating athwartships which are fitted with sets of ultrasonic transducers (up to 50 receivers). In the area echographs, the transducers are switched to avoid picking up spurious echoes. The record is made on a wide paper chart.

1. Introduction

For surveying and supervising water channels, rivers and canals instruments are required which can provide, in one pass along the channel, as much information as possible on the depth of the water, the nature of the bottoms, and the location of obstacles to shipping. The ultrasonic echographs which have been in use for some time in many countries permit only the recording of the depth of water directly under the survey ship, so that during one pass along the channel a longitudinal profile can be recorded, while repeated passes across the channel give a transverse profile. This method is not satisfactory so that in the interests of the safety of shipping, especially in shallow waters, other methods must be sought, which are more suitable for the rapid surveying of waterways. The further development of acoustic echo sounding from linear sounding to area sounding offers new possibilities in this direction.

It is extremely difficult, and only possible within narrow limits, to determine the surroundings of the ship with the aid of directional *horizontal sounding* from the survey ship and to record the results with planimetric accuracy on a paper strip in a clear and intelligible manner. Many disturbing subsidiary echoes from the surface and the bottom, and the relatively short sequence of pulses inherent in the sounding range quickly set a limit to the use of horizontal sounding for the task to be performed here. The position is quite different, however, if we use not horizontal sounding, but only *vertical echo sounding*. A greater sequence of soundings at shallow depths and the elimination of all disturbing echoes make it possible to obtain unambiguous and clear recordings, which are well suited for direct evaluation. But, in conjunction with the profile and area echograph described here, horizontal sounding is usable in rivers and canals to measure the lateral distance from the survey ship to the banks, as will be explained later.

2. Profile Echographs

Transverse profiles can be recorded during a pass along the length of a channel if the survey ship is equipped beneath the hull with a number of transmitters and receivers having a suitable directional effect. This can be provided with lateral outriggers, each similarly equipped with equally spaced transmitting and receiving devices, as shown in Fig. 1. A recording instrument coupled with these multiple measuring points is so constructed that the recording mechanism, pulse transmission triggering device and the stylus can be moved relative to the paper like a typewriter carriage; in passing over a contact strip the instrument scans all measuring points on the outriggers and beneath the survey ship successively. By this means, transverse profile recordings are obtained in a few seconds for profile breadths of up to 50 metres. Figure 2 shows successive profiles of a side channel of the River Weser which have been recorded automatically. Depth is recorded at a scale of 1 : 100 so that 1 cm on the paper represents 1 metre of water depth. If, however, it is desired to record the transverse profiles not one below the other, but continuously side by side on a paper strip, it is possible to employ the conventional type of echograph and to use a slow paper-transport speed for longitudinal profiles and a faster paper-transport speed with simultaneous scanning of the measuring points for transverse profiles. As Fig. 3 shows, this side-by-side representation of transverse profiles is better suited for immediate recognition of when the water depth is more or less than a given value.

A profile echograph of this type has appreciable advantages over the conventional echographs for the recording of transverse profiles while the ship is moving or at a standstill; it remains, however, a linear sounding with which a channel cannot be completely covered, because soundings are made *successively* from the individual measuring points and because the

‡ Laboratory of Dr. Fahrentholz, Echo Sounder Factory, Kiel, Germany.

† This paper has also been published in the *International Hydrographic Review*, January 1963.

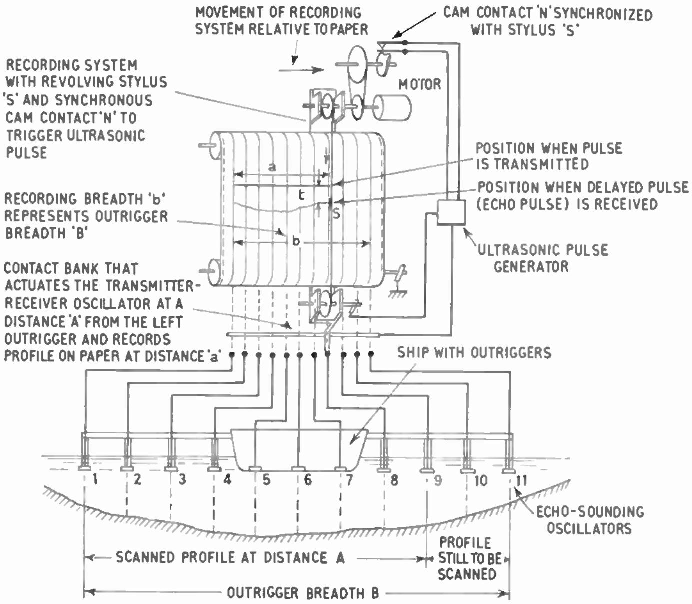


Fig. 1. Principle of the profile echograph.

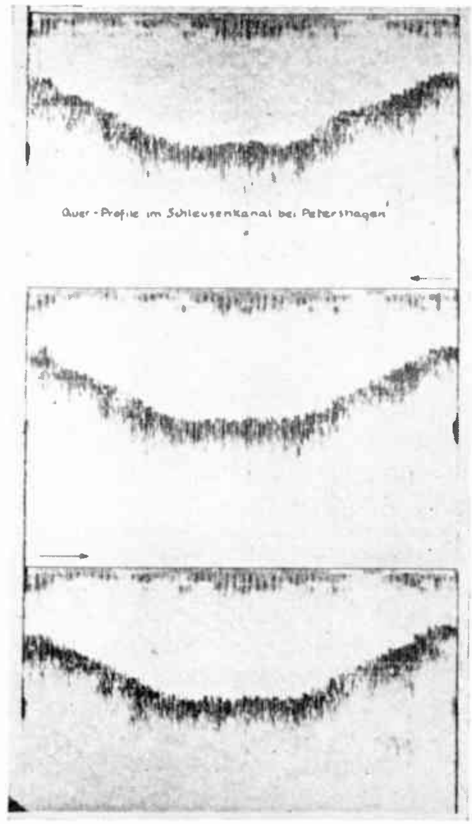


Fig. 2. Canal profiles.

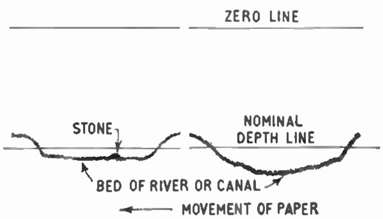


Fig. 3. Transverse profiles side by side on a recording strip.

survey ship has moved on during the recording of a transverse profile.

3. The Area Echograph

If the disadvantages of the profile echograph are to be avoided, an area echograph, as illustrated in Fig. 4, must be built, which transmits *simultaneously* over the entire measuring breadth and receives all the echoes of the various measuring points separately and records them independently on a sufficiently wide paper strip. With this type of apparatus, it is possible to register as many longitudinal profiles at equal intervals from each other as there are measuring points available. The channel bottom is completely scanned, using overlapping index lines; all obstacles protruding above the bottom will be registered and reliably recorded. Figure 5 shows the complete instrument for simultaneous recording of 50 longitudinal profiles one below the other. In the rear closed part of the instrument there is a belt at right angles to the paper, which is fitted with 50 equally spaced styli. The paper is 90 cm wide and moves over the measuring table at a speed that can be varied between 15 and 60 mm/min. The electronic parts of the equipment for transmission of the short ultrasonic pulses and reception of the echoes are housed in the base of the area echograph.

An example of the records obtained by this method is shown in Fig. 6. The spacing between longitudinal profiles is equal on a flat bottom; it varies if the depths of water within the profile breadth covered by the survey ship are unequal. The use of a recording at a depth scale of 1 : 100 on 90 cm wide Teledeltos paper permits the location on a flat bottom of objects larger than about 20–30 cm. Evidence of this performance is given in Fig. 7, which shows a recording made with 10 measuring points in the experimental tank of the laboratory. The walls and floor of the tank are of concrete and are not lined with non-reflective material. The stone lying on a mobile carriage on the floor is moved below the individual measuring points at 3 different speeds and is successively registered by the index lines of the various measuring points and recorded. At low speed, its shape is clearly recognizable. At a higher speed, it appears as a clearly visible point. This ordinary stone is about 35 cm high, 40 cm wide and 50 cm long. In calm water, the instrument measures accurately to within about 4 cm in the

0–20 metre measuring range. The paper feed is set at 15 mm/min.

3.1. *Suppression of Multiple Echoes*

At this point, a few of the difficulties must be mentioned, which were encountered in the development of the area echograph and determined the further course of development of these instruments. With a measuring breadth of 50 m, it is not possible, without making some special provision, simply to combine 50 echo sounders in a single instrument, to transmit pulses with all transmitters simultaneously and record all the echoes picked up by the various receivers in clear and proper relation to each other so that well-defined, interrelated longitudinal profiles are obtained. Each transmitter sends out radiations of more or less large amplitude to several neighbouring receivers; in an experimental tank of concrete innumerable echoes were registered during the entire sounding period. This is a difficulty which also occurs especially in shallow waters and near embankments. It was overcome by equipping each receiver amplifier with a special electronic switch which permits only reception of the first echo following the transmission of the sounding pulse. All multiple echoes are suppressed and sorted so that a clear record is obtained. Use is also made of electronic devices which, if no echo occurs, produce artificial echoes in the equipment in order to make all amplifiers ready for reception again a short time after transmission, and prevent registration of all transmitted pulses. This automatic suppression of multiple echoes, which is a precondition for the construction of an area echograph, also permits the simultaneous setting of all amplification stages of the various reception channels with a single control knob. Uniformity of amplifiers, transmitters and receivers is not imperative for the response of the automatic pulse sorting system; a certain minimum amplification is merely required that is dependent on water depth and the nature of the bottom. Even small objects which give very weak echoes can be recorded when lying on concrete with an echo one thousand times stronger. As could be seen in Fig. 7, the recording shows the outline of the stone. So it cannot simply be discerned from the recording whether the protuberance on the bottom is a stone or a mound of earth.

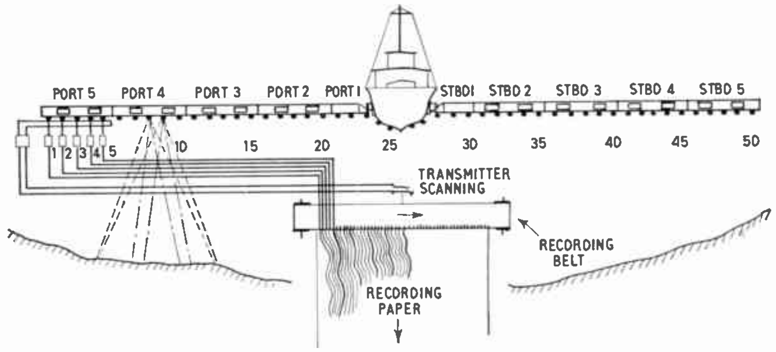


Fig. 4. Principle of area echograph showing the arrangement of the 50 measuring points on the outriggers and beneath the survey ship.

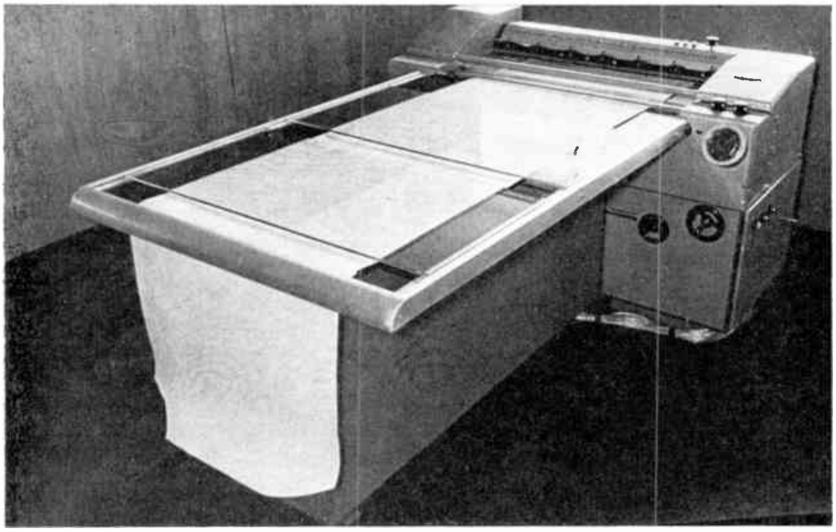


Fig. 5. Area echograph.

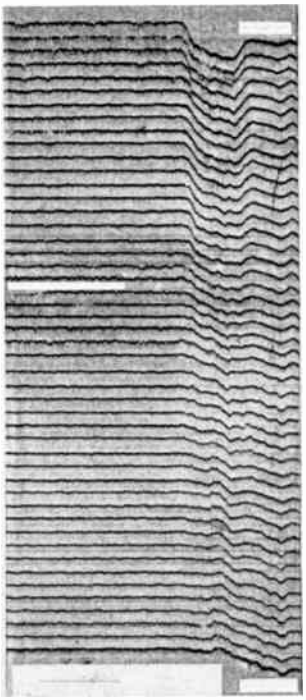


Fig. 6. Area echograph record.

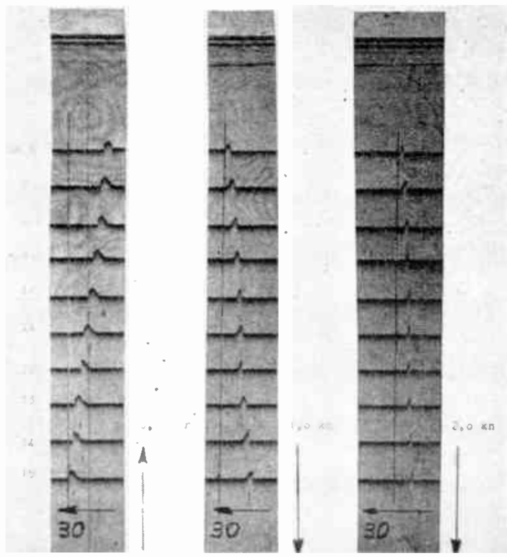


Fig. 7. Area echograph recording of a stone in an experimental tank.

This is a disadvantage which is caused by the fact that it is not the echo itself which is recorded directly on the paper, but a short electric pulse produced by triggering a multivibrator. Therefore, the further development of the area echograph led to a true-to-life recording in which short and long echoes or small and large echoes are registered as nature produces them. Figure 8 is an illustration of such a recording, in which fish can be distinguished from the ground and from stones.

3.2. *Recording of Profiles*

The use of the area echograph in practical applications has further shown that in surveying a water channel it is not always necessary to record as many longitudinal profiles as there are measuring points on the survey ship and the outriggers. The results from several neighbouring measuring points can always be consolidated, as shown in Fig. 9, and recorded as a single longitudinal profile, giving only 6 or 12 longitudinal profiles for a measuring breadth of 48 metres. The representation on paper is clearer and the channel is covered without any gaps despite this simplification.

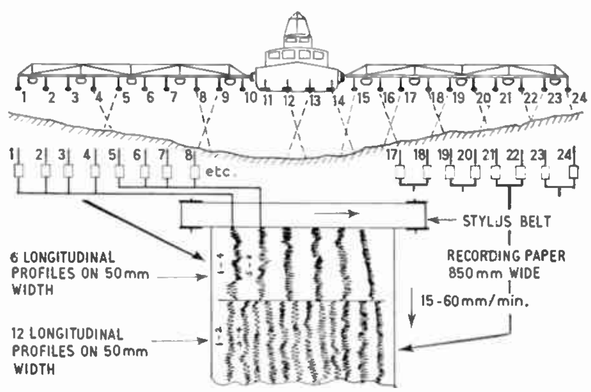


Fig. 9. Area echograph with 6 or 12 or more longitudinal profiles.

All the area echographs described record only longitudinal profiles one below the other. If transverse profiles are required, they can be drawn simply by superimposing transparent paper and using compasses. However, a special supplementary transverse profile echograph is also available, as illustrated in Fig. 10. When a button is depressed, it records a

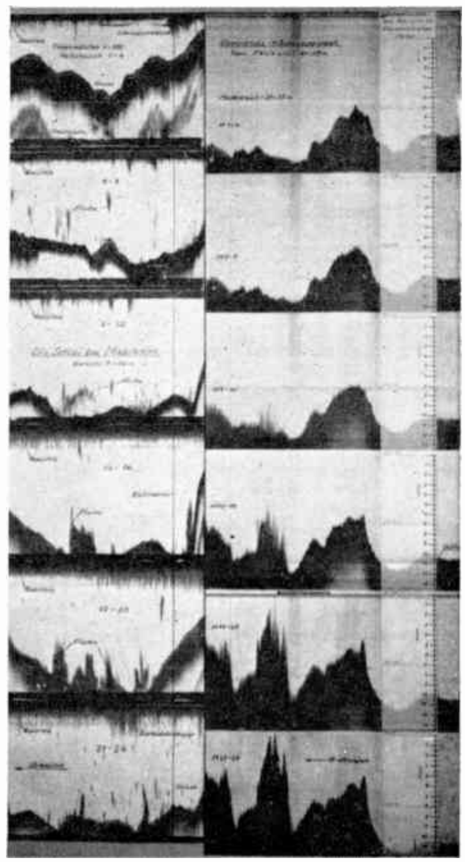


Fig. 8. Area echograph recordings.

transverse profile in $\frac{1}{4}$ second. In this instrument a number of styli mounted side by side simultaneously record the water depth of the transverse profile in a single pass over the paper. Time marks on the area echograph and the transverse profile echograph guarantee proper co-ordination.

If the distance to the banks is to be recorded continuously as the ship moves along the channel, in order to determine the course followed, sharply focused horizontal sounding can be used. Figure 11 shows recordings of a stabilized bank and a reed covered bank of the Weser. The groynes under the water are also clearly recognizable.

Fig. 10. Rotograph for recording transverse profiles in 0·2 seconds.

3.3. *Transmitting and Receiving Transducers*

Profile and area echographs have hitherto operated with an ultrasonic frequency of 40 kc/s and used magnetostrictive transmitters and receivers of pure nickel. Nowadays, the latest instruments use 100 kc/s and have transmitters and receivers of barium titanate. The measuring points below the survey ship and on the outriggers are arranged to measure water depths of less than 5 m at intervals of 1 m at right angles to the sailing course. In the case of instruments for greater depths, the spacing is 2 m. The directional effect of transmitters and receivers is such that at least two measuring points overlap so that the bottom is completely covered. Older types of instruments made soundings with 2 pulses/second, modern instruments use 15 pulses/second and more.

4. Survey Ships

A few words should be said about the development of the survey ships and outriggers which are necessary for the employment of the profile and area echographs.

The techniques of recording longitudinal and transverse profiles in water channels, using a large number of vertical transmitters and receivers, requires the design of outriggers serving to extend the breadth of the survey ship. The idea of sailing through narrow channels with outriggers was initially rejected by sailors; however, after a 10-year development period, it has been proved that sailing with outriggers in rivers and canals involves no difficulty, and it has been found that a ship with outriggers has good manoeuvrability, provided that it has two screws. In all cases, however, it must be possible to retract the outriggers simply and quickly while the ship is under way, in order to make room for oncoming vessels in narrow channels. In the beginning, therefore, development work and practical tests were carried out on outriggers for survey work on quiet rivers and canals—not for the open sea—which could be run out to a length of up to 25 m on each side of the ship and which could be folded up and then folded back towards the stern.

The outriggers so far designed for river survey ships have not proved stable enough for the open sea, and extension and retraction with hand winches is too troublesome and requires an operating crew of several men. Moreover, for ships that are to carry out survey work with area echographs off the coast, it is necessary to be able to take the outriggers on board simply and quickly.

To satisfy these requirements, development work has culminated in the design of an all-purpose survey ship, a model of which has been built and is illustrated in Fig. 12. The outriggers consist of stable fire-escape type extending ladders, over which crew members can move safely. They can be run out to the side and each rests on a float. Both outriggers can pivot about a horizontal axis, to compensate for

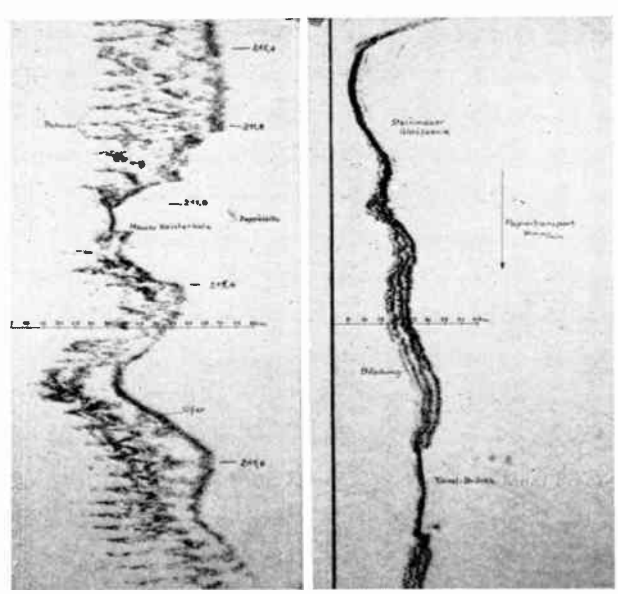
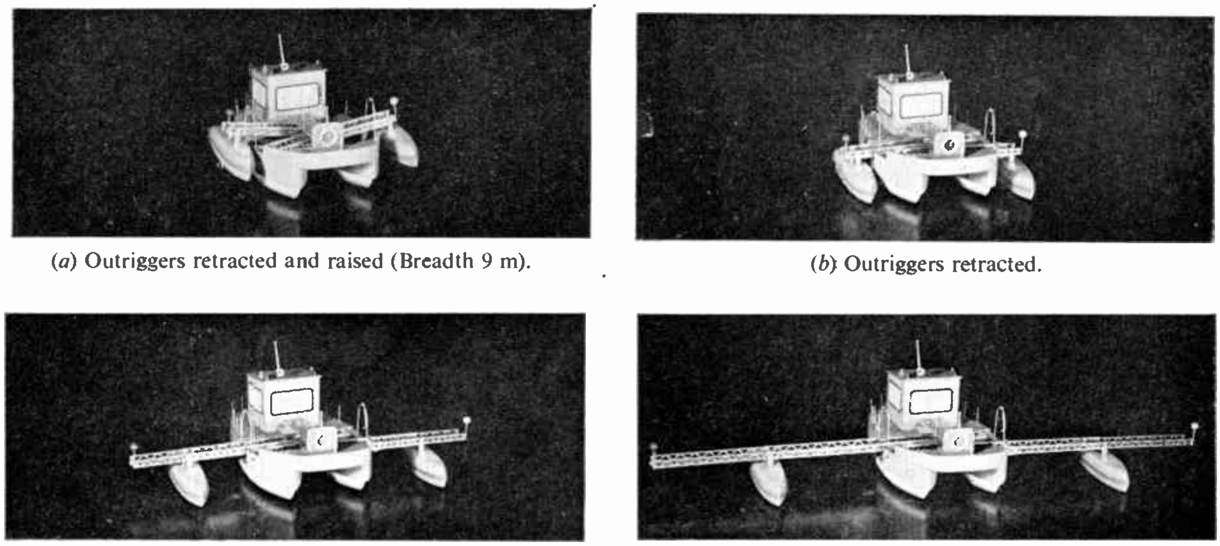


Fig. 11. Horizontal sounding on the river Weser.
(*Left*) Representation of reed-covered bank with groynes.
(*Right*) Representation of a canal embankment with intervening bridge pier.



(*a*) Outriggers retracted and raised (Breadth 9 m).

(*b*) Outriggers retracted.

(*c*) Outriggers half run-out.

(*d*) Outriggers fully run-out.

Fig. 12. Model of a sonar survey ship with outriggers of variable span.

ship movements in heavy seas. The bearing of the port outrigger is on the starboard side of the ship, and that of the starboard outrigger on the port side. Model experiments have shown this arrangement to be particularly favourable. The outriggers are run out and retracted electrically requiring no manual labour. In the retracted position, they can be hydraulically raised out of the water, so that the survey ship can reach the harbour safely in bad weather. Survey ships of this type can be built for variable measuring breadths of 7–30 m, 9–40 m and 10–50 m. It is certain that these vessels are highly suitable for rivers and canals, but they still have to be tested in the open sea.

5. Conclusions

A description has been given of the new profile and area echographs and the respective survey ships and outriggers, outlining the development from linear acoustic sounding to area sounding. The area echographs customarily used at the present time permit frequent, rapid and complete checking of the depth of navigation channels in rivers and canals and the rapid elimination of obstacles. They also make it possible to plan and carry out dredging work accurately, so that their use contributes substantially towards increasing the safety of shipping. The development of instruments and survey ships for operation on the open sea is not yet concluded.

6. Acknowledgments

The work was carried out in the author's laboratory and was furthered by research contracts for the German Waterways and Shipping Authority and the Kungl. Sjöfartsstyrelsen in Stockholm.

Manuscript received by the Institution on 4th May 1962. (Paper No. 842/SS8.)

CONTRIBUTION TO THE DISCUSSION

Unfortunately Dr. Fahrentholz was unable to be in Birmingham to present his paper in person. Its details were summarized by the Chairman and a general discussion on sonar surveys took place. The major contribution is printed below.

Dr. H. Maass†: In waterways survey work during recent years the main problem has not been to obtain sufficient information from different systems of recordings of echo sounders, but to evaluate the material rapidly. This led to looking for a method by which one may obtain the information at first glance in a map-like presentation.

In the past for river and channel survey work echo sounding methods have been in use by which cross-sections of the river respectively the channel were recorded in intervals on the recording paper. Besides this the system in Dr. Fahrentholz's paper provides simultaneously a great number of longitudinal recordings. For this purpose a straight line of a great number of transducers with relatively sharp beam pattern to the bottom was used, equally distributed over, for instance, a length of 40 metres.

† Atlas-Werke A.G., Bremen.

The centre part is installed in the flat bottom of the ship, the left and the right parts are installed on removable beams of wood or steel structure with supporting streamlined bodies at the end and maybe in the middle.

In the new bottom chart recorder equipment the same arrangement of transducers is being used. In order to obtain a map-like presentation of the search area the system works as follows:

A series of pulses is transmitted from left to right in a fast sequence, and the received echo is fed to a system of cams arranged at right angles to the paper speed of the recorder. This paper speed will be controlled by a sound wave Doppler-Log according to the ship's actual speed. The scales are chosen so that one obtains a map-like presentation.

From different depths the echo will be received at proportionally different times after the transmission of the individual pulse. An electronic circuit provides that the intensity of recording is inversely proportional to the elapsed time such that echoes appearing first and belonging to a minimum depth are recorded in dark black, and echoes arriving later are recorded in four steps either with a fainter black, dark grey, light grey or white—which means no recording at all. The steps of time-interval can be either 1 m or 50 cm or 10 cm, and the centre depth of recording can be set to any value between 2 m and, for instance, 16 m.

By this method one obtains therefore a map-like presentation of the bottom from different horizontal levels similar to the different colours of a geographic map.

The method has in the meantime been improved by arranging a second recording system to work alongside the first one and in parallel in which the centre depth set to the intended value is identical with the other recording, but where the steps are, for instance, 10 cm only, whereas the other recording shows steps of 1 m. A third recorder can provide cross-sections in the old way using the same echo information, but stored in a capacitor system and scanned at random by pushing a button at larger time-intervals, for instance half a minute. Building up this cross-section takes about 10 seconds.

Echo sounders using a bottom chart recorder have been used satisfactorily on board two large sea-going dredgers.

Dr. S. Fahrentholz (*in reply*): The bottom chart recorder described by Dr. Maass is capable only of providing a relatively rough survey of the river bed, as the blackening of the chart can only be made sufficiently distinct in the three tones light grey, dark grey, and black. The resulting accuracy of measurement does not provide sufficient accurate information on the depth of water and the size of obstacles to shipping; in respect of the accuracy of measurement the bottom mapping recorder is, therefore, not comparable with the area echographs I described. Besides, the 'continuous simultaneous sounding' of the area echographs with all transducers on the whole width of measurement is superior to the 'successive sounding' of the bottom chart recorder, as a high sounding recurrence frequency together with the higher accuracy of measurement at a given speed is the prerequisite for the gapless mapping of the traversed bed and all obstacles. The experience with 14 profile and area echographs on board survey ships and dredgers has proved that the presentation of longitudinal profiles achieves the advantages claimed in my paper.†

† See also paper by Bernhard Möhlmann, in *Die Bautechnik*, 5, pp. 154–60, 1963.

Radio Engineering Overseas . . .

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

LOGARITHMIC PERIODIC ARRAYS

Logarithmically periodic arrays have pattern and impedance characteristics which are essentially independent of frequency over a large range, i.e. they are broadband structures. An engineer at the University of Melbourne, describing a research programme on this type of aerial points out that it is still not clear why the aerial should radiate in the direction of the apex, or feed point, whereas other aerials similar in shape to the log. periodic aerial, such as the horn, radiate towards the open end of the aerial. Since it is the current distribution which determines the direction and shape of the far-field radiation pattern, current distribution measurements were made on a log. periodic aerial. These distributions clearly show that an active region exists which remains at a fixed electrical distance from the apex. The current is negligible in the conductors of length greater than one-half wavelength. The measurements have also shown the effect of frequency on the current distribution and have verified the fact for an aerial to possess characteristics which approach that of frequency independent aerial, it is necessary that the end-effect is negligible, and that the radiation pattern is due to a finite active region which remains identical and at a fixed electrical distance from the origin or apex of the aerial for all frequencies over the designed bandwidth.

"The log. periodic as a broadband aerial", B. MacA. Thomas. *Proceedings of the Institution of Radio Engineers Australia*, **24**, pp. 355–63, April 1963.

DATA TRANSMISSION

The February 1963 issue of *L'Onde Electrique*, journal of the Société Française des Électroniciens et des Radioélectriciens, contains 16 papers on various aspects of data transmission, including descriptions of systems at present in use or under development in France. The titles of some of the papers are:

"A numerical data transmission system", M. Coiron, G. Dupire and B. Lorimy, pp. 141–47.

"Differential modulation data transmission system using phase inversion", J. Claisse, pp. 148–52.

"The 'Marathon' system of teletransmission. Principles and applications", J. Fuzellier, pp. 170–76.

"Results of data transmission tests over telephone links", J. Labeyrie, pp. 94–105.

"Transmission of numerical information", M. Poliet, pp. 106–16.

"Codes for providing protection against errors in data transmission", F. Corr and E. Gorog, pp. 117–27.

"Cyclic codes and data transmission", S. Fontanes, pp. 128–40.

"An application of data transmission in the field of telecommunications: centralized taxation and accounting in telephone exchanges", R. Legare, J. Dondoux and J. C. Lavenir, pp. 214–19.

"Mechanography by wire", L. Durand, pp. 220–30.

"Data treatment using numerical transmission over long distances", A. Desblache, pp. 243–50.

"Phase inverting modulator/demodulator for the transmission of data on telephonic circuits", J. M. Pierret, pp. 177–85.

"'The Rapidata S' synchronous system for the high speed transmission of intelligence", A. Girinsky and P. Roussel, pp. 186–98.

"High speed data transmission", H. Cohn, pp. 199–206.

"Tests on high speed transmission of coded information over an experimental link between France and Holland", J. Touchard, pp. 207–13.

L'Onde Electrique, **43**, No. 431, February 1963.

INFINITELY VARIABLE FERRITE PHASE SHIFTER

Work has been carried out in Japan on some new types of phase shifter which have continuously and infinitely variable phase shift and quick response. These are employed to perform high speed continuous beam scanning and receive microwave power from several antennae in a space diversity system and ensure that it is always in phase. It is shown from experimental results that the phase shifts can be varied continuously with the scanning speed of 18 000 deg/s. Amplitude variation during the phase shift period and the phase discontinuity of the microwave power during the switching time are less than 0·3 dB and 5 degrees respectively.

"Infinite variable ferrite phase shifter", T. Kitsuregawa, S. Nakahara and T. Kondo. *Mitsubishi Denki Laboratory Reports*, **4**, pp. 15–30. January 1963. (In English.)

TRANSISTOR NOISE

The r.f. noise properties of transistors, that is to say thermal and shot noise effects at frequencies exceeding 1/1000 of the cut-off frequency, can be expressed in terms of the electric parameters of the transistor, and the results are presented in a simple and universal form in a recent Dutch paper. From these results all quantities related to noise can be derived, using the minimum of calculations, provided the current gain, the base resistance and the biasing point with the corresponding cut-off frequency of the transistor are known.

"Noise properties of transistors at high frequency", W. Smulders. *Electronic Applications*, **23**, No. 1, pp. 1–25, 1962–63. (In English.)