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 NOVEMBER 1963 NUMBER 5

SPECIAL GENERAL MEETING

NOTICE IS HEREBY GIVEN that a Special General Meeting of the Corporate Members of the Institution will be held at the London School of Hygiene and Tropical Medicine, Keppel Street, Gower Street, London, W.C.1, on Wednesday, 27th November 1963, at 5.45 p.m., in accordance with Article 20 of the Institution's Charter for the purpose of considering and if thought fit confirming the following Resolution passed at a meeting of the Council of the Institution held on Thursday, 10th October 1963:

RESOLUTION OF THE COUNCIL

That the name of the Institution be changed to The Institution of Electronic and Radio Engineers and that the Charter of the Institution be altered accordingly in the following manner:

1. By deleting the words "The British Institution of Radio Engineers" where these words occur on the second occasion in Article 1 of the Charter and substituting therefor the words "The Institution of Electronic and Radio Engineers".

2. By deleting Article 13 of the Charter and substituting therefor the following new Article:

   "13. A Member (including an Honorary Member who, when elected an Honorary Member, was already a Member) shall be entitled to the use after his name of the designation "M.I.E.R.E.": and an Associate Member the designation "A.M.I.E.R.E."."

3. By deleting Article 14 of the Charter and substituting therefor the following new Article:

   "14. Every person being at any time a Corporate Member of the Institution may so long as he shall be a Corporate Member take and use the name or title of or describe himself as a Chartered Electronic and Radio Engineer."

The following Resolution will be put to the meeting:

RESOLUTION

That the Resolution of the Council set out in the notice convening this meeting be and the same is hereby confirmed.

(continued overleaf)

Subject to the above Resolution of the Council being confirmed to consider and if thought fit pass with or without amendment the following Resolution in accordance with Article 16 of the Institution's Charter:

RESOLUTION

That Bye-laws 1 and 9 of the Institution be amended as follows:

1. By deleting the words "The British Institution of Radio Engineers" in the first sentence of Bye-law 1 and substituting therefor the words "The Institution of Electronic and Radio Engineers".

2. By deleting the existing Bye-law 9 and substituting therefor the following new Bye-law:

"9. Each Member and each Honorary Member who when elected an Honorary Member was already a Member is entitled to use after his name the designation "M.I.E.R.E." and to describe himself as a Chartered Electronic and Radio Engineer. Each Associate Member is entitled to use after his name the designation "A.M.I.E.R.E." and to describe himself as a Chartered Electronic and Radio Engineer.

A Corporate Member practising—

(i) under the title of, or as an officer or employee of a limited company or,

(ii) in partnership with any person who is not a Corporate Member of the Institution under the title of a firm

shall not use or permit to be used after the title of any such company or firm the designation "Chartered Electronic and Radio Engineer" or "Chartered Electronic and Radio Engineers" or describe or permit the description of such company or firm in any way as "Chartered Electronic and Radio Engineer" or "Chartered Electronic and Radio Engineers"."

By Order of the Council,

31st October, 1963

GRAHAM D. CLIFFORD,
Secretary

(Members who are unable to attend the Special General Meeting are invited to use the form of proxy enclosed which should be returned to the Institution not later than forty-eight hours before the time for holding the meeting. The person appointed to act as proxy must be a Corporate Member of the Institution, and the President or the Secretary will be pleased to act in this capacity.)

The above Notice was posted to all Corporate Members in Great Britain on 31st October 1963.

U.D.C. 621.382.1: 681.32

Tunnel Diode Storage System with Non-Destructive Read-Out

By

D. B. G. EDWARDS,
Ph.D., M.Sc.†

AND

M. J. LANIGAN,
Ph.D., M.Sc.‡

Presented at a Computer Group Symposium on "Tunnel Diodes as Switching and Storage Elements" in London on 23rd January 1963.

Summary: A brief survey of some tunnel diode storage systems is presented and a description of a particular system with a non-destructive read-out feature is given. The design of this system is discussed in some detail, and an assessment of the operating margins is included. In a unit of 128 words of 24 digits per word a read access time of 150 mμs is indicated with a read cycle time of 250 mμs. The write cycle time is 330 mμs.

1. Introduction

In the *Atlas* computer at Manchester University the present B-store is a fast core store of 128 twenty-four bit words operated in a word selected two-core-per-bit partial flux switching mode.[1] The read access time of this store is 0·35 μs, with a cycle time of 0·7 μs, and this performance limits the speed of a number of important machine orders.

The primary object of the work described in this paper is the design of a store of the same capacity as the core B-store with an improved performance and comparable cost. In this context a read access time of 0·15 to 0·25 μs is acceptable with a cycle time not greater than 0·4 μs. The Esaki[2] or tunnel diode offers an attractive alternative to square-loop magnetic cores for stores of about this capacity since the properties of digital storage, selection and fast switching can be realized at relatively low power levels. In such a store the operating speed is limited principally by the physical size of the store, the address selection and read-write mechanism, as the switching speed of the storage elements is extremely fast, of the order of 1 to 10 mμs.

In a tunnel diode store there is also the possibility of non-destructive reading of the stored information, which allows the read cycle time to be less than the write cycle time. This feature is of interest because, on the average, there are rather more read operations from the B-store than write operations to it, the ratio varying typically between 6 : 1 and 2 : 1.

The *Atlas* machine also contains a number of smaller random access stores, for example the tape buffer and check-sum stores. These are at present realized in terms of magnetic cores and transistor staticizors respectively. These units might be replaced by tunnel diode stores with an economic advantage.

The work described is also of more general interest in that the techniques developed could be applied to the next generation of machines which might, for example, have a fast tunnel diode main store backed up by a core store, the combination being operated as a one level system similar to that employed in *Atlas* involving the main core and drum stores.[3, 4]

The main alternatives to tunnel diodes for high speed storage applications are thin magnetic films and super-conducting devices. The practical difficulties associated with these developments are quite severe and present progress in these fields is slow. In practice a magnetic film store[5] is not likely to be much faster than a tunnel diode store of the same capacity, while the operating power level of the film store is much higher and the output signals are at a lower level.

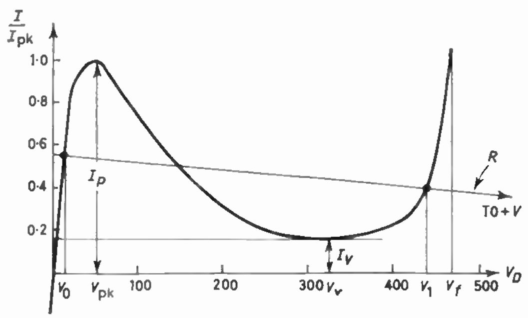
2. Tunnel Diode Characteristics

The typical characteristic of a germanium tunnel diode is illustrated in Fig. 1(*a*). The high positive conductance of the device at low reverse and forward voltages is due to the 'tunnelling' effect. The tunnelling current is a decreasing function of the forward voltage when this exceeds a critical value and the second region of positive conductance is due to the normal rectifying action of the p-n junction. In the range $V_{pk} \rightarrow V_v$ there is a region of negative conductance as the tunnel current decreases more rapidly than the rectifier current increases. A simple storage element, (Fig. 1(*b*)) is then obtained by superposition of a suitable load line, R, in Fig. 1(*a*) which gives the two stable potentials V_0 and V_1.

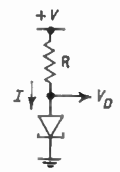
At room temperatures a feature of the tunnel diode is the precision of the characteristic. This is most marked in the case of the peak current, I_{pk}, which is typically defined to within $\pm 5\%$; similarly V_f may be controlled to limit of $\pm$ 10%. The valley characteristics, V_v and I_v, are not so well defined and an

† Electrical Engineering Department, University of Manchester.

‡ Formerly in the Electrical Engineering Department, University of Manchester; now with The Plessey Co. (U.K.) Ltd.



(*a*) D.C. characteristic and load line.



(*b*) Simple storage element.

Fig. 1. Tunnel diode characteristics.

I_{pk}/I_v ratio of 5 to 10 is typical of germanium units with a corresponding variation in the valley voltage of about $\pm$ 20%. For devices in a given material the voltages V_{pk}, V_v and V_f are virtually independent of the peak current. Thus the positive and negative incremental impedances at a given voltage are approximately inversely proportional to the peak current and for high current devices these can be as low as 1 or 2 ohms. This characteristic is useful in certain applications where the diode is used as a voltage definition element.

2.1. *Temperature Effects*

The tunnel diode characteristic is sensitive to temperature variations, but these do not present a serious problem in storage applications where control of the ambient temperature is possible and the device dissipation is only a few milliwatts. The figures quoted in Table 1 apply to the particular germanium units used in the development of the system described in Section 3.2.

Table 1†

Temperature co-efficients in the range -40 to $+70°$ C

Parameter	Temperature Co-efficient (% change per deg C)
I_{pk}	$\pm$ 0·3
I_v	+ 0·5
V_{pk}	− 0·4
V_v	− 0·4
V_f	− 0·25

† Communicated by J. Harwood.

The temperature co-efficient of the peak current may be either positive or negative, an effect which is not yet fully understood.

2.2. *Switching Characteristics*

The switching speed of the element is limited by the junction capacity and circuit conditions. An approximate analysis of the problem, which assumes a constant junction capacitance and makes a straight line approximation to the diode characteristic is presented in Appendix 1. The results obtained are in fair agreement with practical observations and indicate the limitations imposed on the switching speed by the operating conditions.

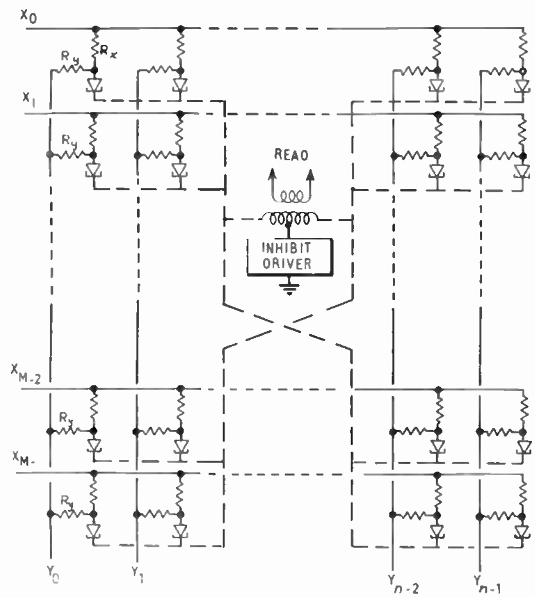
3. Tunnel Diode Storage Systems

The possible arrangements of a storage system to give parallel read-out of a number of digits fall into two broad classes, matrix and word selection systems. In the first of these, a matrix plane is assigned to each digit position of the word and these planes are driven in parallel. Selection is achieved by a coincident voltage technique which employs the current discriminating properties of the storage elements. With this technique the final stage of address decoding occurs at the selected digit; this is not so in a word selected store where the digits of a given word are arranged as a unit which is selected when the complete address has been decoded.

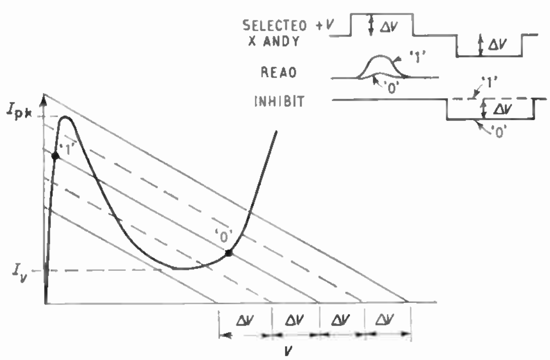
3.1. *A Matrix System*

A simple matrix system, similar to that considered by Rajchman[6] is illustrated in Fig. 2(*a*). The resistive load line defined by V and $R/2$ is chosen to give a 2 : 1 selection ratio at the peak and valley current limits, Fig. 2(*b*). The reading operation is destructive and the occurrence of a significant current change in the selected element indicates the presence of a '1' digit. In an $N \times M$ matrix there are $(N-1)+(M-1)$ elements half-selected during the read operation which generates interference signals in the common read wire. The effect can be minimized by a cancellation technique such as that illustrated in Fig. 2. Complete cancellation is not possible under all conditions because the magnitude of the 'disturb' signals is a function of the state of the half-selected digits. In this system writing is achieved by means of an inhibit technique which prevents the automatic set to '1' in the write phase in those digit positions where the storage of a '0' digit is required.

The main disadvantage of this and similar systems is that the X and Y drivers must be able to supply pulses of relatively high power. The current is high because of the large number of elements in parallel (nN or nM for a store $N \times M$ words of n digits/word), and the voltage is high so as to give adequate current definition. Typically several volts and some hundreds



(*a*) Layout of wiring.



(*b*) Load lines during selection.

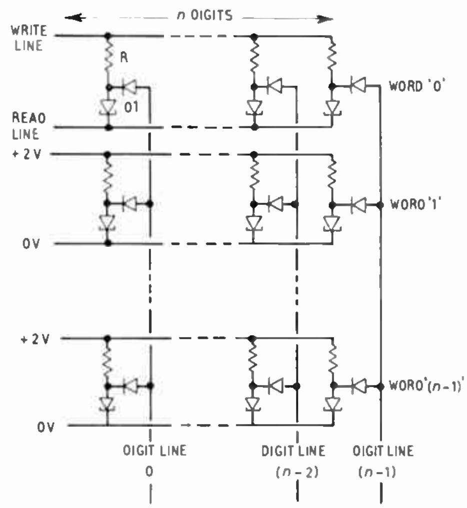
Fig. 2. A matrix system.

of milliamperes must be provided. For these reasons a word selected arrangement is preferred. The system described can easily be transformed to a word selected arrangement in which the X drivers become word selectors and the Y drivers control the digit lines which are then associated with individual read amplifiers. This effects an immediate reduction in the word driving power but the digit driver must still supply a large number of parallel elements and the resistor R_y in Fig. 2(*a*) may be replaced by a conventional diode in a word selected arrangement. This destroys the symmetry of the storage element and the writing process then becomes a two phase operation.

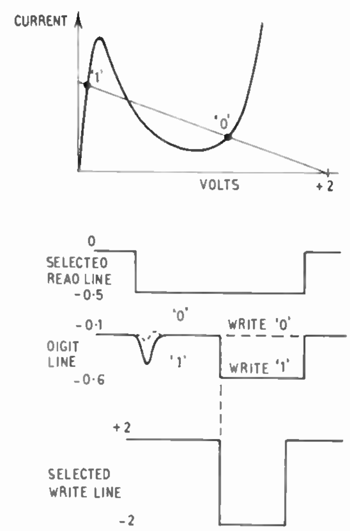
3.2. *Word Selected Systems*

One possible arrangement[7] of a word selected system is shown in Fig. 3(*a*). To read the selected word the read line is switched to a defined negative level which causes the read diode (D1 etc.) to conduct only if the tunnel diode is in the low voltage state, Fig. 3(*b*). The reading operation is destructive, and the read phase is followed by a period in which the bias voltage between the read and write lines is reversed. This sets all the storage diodes in the selected word to the '1' state, and writing to the '0' state is then achieved when the bias voltage is returned to its normal level by causing the read diodes to conduct again in the appropriate digit positions.

A feature of this system is that the read pulse is of short duration, for example 1 mA for 5 mμs, defined

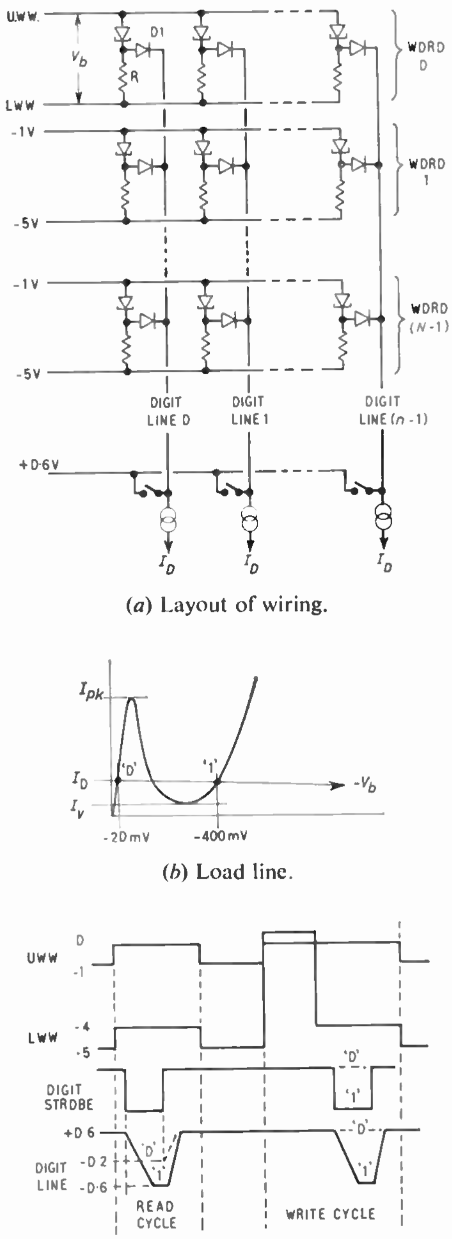


(*a*) Layout of wiring.



(*b*) Load line and waveforms.

Fig. 3. A word selected system.



(*a*) Layout of wiring.

(*b*) Load line.

(*c*) Waveforms.

Fig. 4. System for non-destructive read.

by the switching time of the tunnel diodes and the characteristics of the read diodes. The read pulse duration is then comparable with the delay in the digit line in a large store and requires considerable amplification to operate conventional logic circuits.

It is difficult to achieve a non-destructive read with this system, since the tunnel diode current during the reading process depends critically on the magnitude of the read line voltage, the read diode impedance and the characteristics of the digit wire, which is likely to have a low capacitive reactance because of the depletion capacitance of the unselected read diodes. Non-destructive reading is possible with such a system[8] by designing the digit line as a matched transmission line of fairly low impedance (100 ohms) and close control of the selected read wire voltage levels.

3.2.1. A word selected system with non-destructive read

A schematic diagram of this system[10, 11] is shown in Fig. 4(*a*). A constant bias voltage, V_b, is normally maintained between the upper and lower word wires of each word and a virtually constant bias current is used, Fig. 4(*b*). In the quiescent state all the read diodes (D1 etc.) are reverse biased with the switches, S, in the closed position and supplying the digit currents I_D. In a read cycle the levels of the selected word wires are each increased by 1 V and the digit strobe is operated which opens the switches (Fig. 4(*c*)). The digit current I_D then flows in the digit wires which fall in potential until the read diodes of the selected word are carrying I_D. The magnitude of I_D is less than $(I_{pk}-I_0)$ so that the stored information is not destroyed by the passage of this current (Fig. 4(*b*)). The state of a stored digit is indicated by the final level of the digit line, the difference signal which is available for a controlled time, being approximately 0·4 V.

The first part of a write cycle is used to 'clear' the selected word to zero by reversing the direction of the bias voltage, so causing a small reverse current to flow in the selected storage elements. The bias voltage is then restored and the word wires left at the selected levels. During this time the value of I_D in those digit positions, where writing to a '1', is required is made greater than $(I_{pk}-I_0)$, and this then occurs when the digit strobe is operated.

3.2.2. Factors limiting the speed of the system

When the address has been decoded the duration of the store cycle is limited by two main factors:

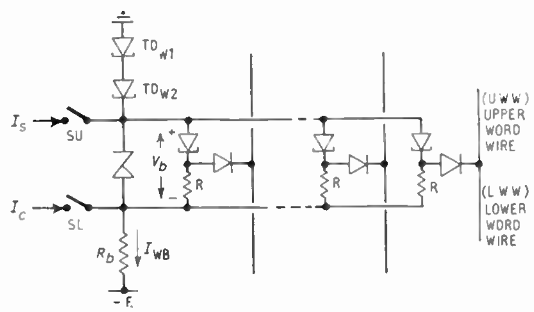
(*a*) the time required to change the quiescent voltage levels of the word wires to and from the selected levels and

(*b*) the rates at which the digit wire voltage can change.

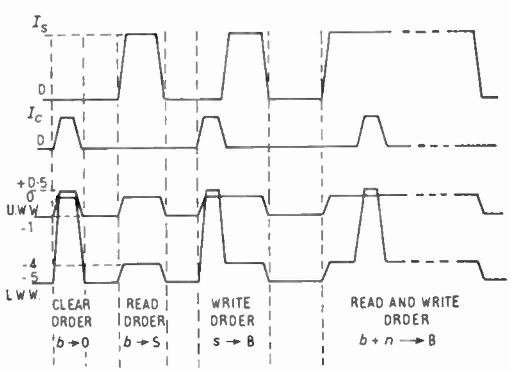
The two factors are not unrelated since the first resolves into a current switching problem, in which the magnitude of the switched currents is an increasing function of I_{pk}, the peak current of the storage diodes. Similarly the rate at which the digit line voltage can change in the read cycle is limited by the magnitude of the read current, which must be less than $(I_{pk}-I_0)$, and the distributed capacitance of the digit line. To minimize this second factor, $(I_{pk}-I_0)$ should be large. This can most easily be achieved by choosing a large value for I_{pk} and operating the storage elements at an approximately constant bias current which is somewhat greater than the valley current, I_v. Here a comparatively large bias voltage is used, which is maintained even when the word is selected for a read or write operation by moving both word wires in unison. In practice a compromise between factors (*a*) and (*b*) is made, which is assisted by splitting the digit line into four sections to reduce capacitance effects.

3.2.3. Definition of the word wire levels

The technique adopted for defining and switching the levels of the upper and lower word wires is illustrated in Fig. 5(*a*). The word bias voltage, V_b, is provided by a low voltage Zener diode, which is normally maintained in the avalanche condition by a component of the current I_{WB}. When the switches SU and SL are open this current flows in the selection tunnel diodes, TD_{W1} and TD_{W2}, which are set in a high voltage high current state. These conditions represent the unselected state of the upper and lower word wires.



(*a*) Bias voltage and selection diodes.



(*b*) Waveforms

(*b*) B = (contents of), selected store line.

(*s*) S = (contents of), another storage register.

n = a number.

Fig. 5. Definition of word wire levels.

Selection of the word is achieved by closing the switches SU and SL in a sequence which depends on the type of store cycle required. In the 'clear' order (Fig. 5(*b*)) SL is closed and a current I_C, which is of sufficient amplitude to cause the Zener diode to switch to the forward conducting state, flows into the lower word wire. For a 'read' order the upper word wire is selected by closing the switch SU so that the selection tunnel diodes are switched to a reverse current condition by the action of I_S. The write operation is obtained by closing SL and SU in turn and a 'read-write' order is also available in which the contents b of the selected line are read out to the arithmetic unit and manipulated therein before the answer, for example $b+$ a number, n, is written back to the same store line B.

The use of a Zener diode to provide the word bias voltage allows this voltage to be fairly large and well defined. The switching speed of the Zener diode to and from the high voltage condition is a function of the current available for switching a 'charge constant', ΔQ_Z, associated with the diode. Transition times of the order of 20 to 30 mμs can be obtained in practice. This can be improved by replacing the Zener diode with a number of series connected tunnel diodes, but the relatively slow switching of the Zener diode is unimportant because a change of state is only required during the 'clear' and 'write' orders.

4. System Design

The practical design of the system described in Section 3.2 is considered under three main headings:

(1) The choice of the storage element bias resistor and system operating currents.

(2) The word selection mechanism.

(3) The digit circuitry.

A schematic layout of the complete system is given in Fig. 6. The store is divided into four pages each of 32 words so that the digit wire assigned to a given digit is in four sections. Each section is associated with a read/write switch which is operated by the appropriate version of the digit strobe. The outputs of the four read amplifiers of a digit are combined and feed a common discriminating and output staticizor.

Attention has been concentrated on a design using 5 mA peak current diodes in the storage elements. These represent a fair compromise between read-out speed and high operating currents. Some design calculations which illustrate this point are given in Appendix 2.

4.1. *Choice of Storage Element Bias Resistor and Operating Currents*

The storage element bias resistor is chosen to give the maximum possible value to the quantity $(I_{pk}-I_0)$,

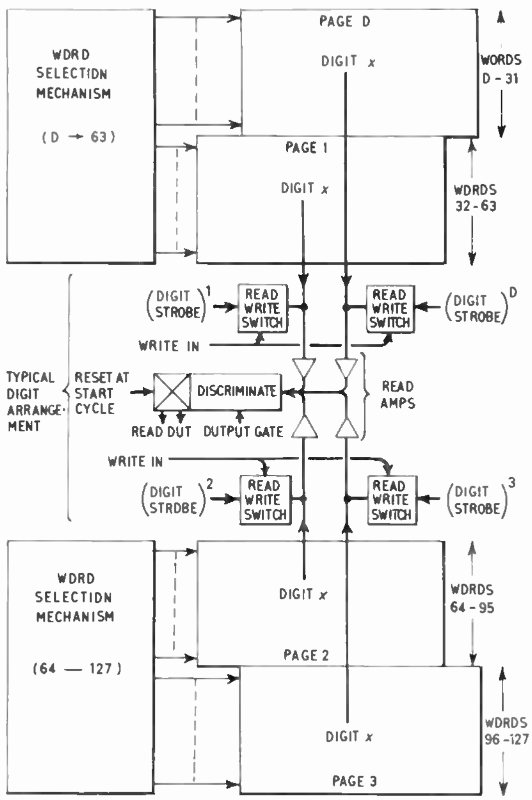


Fig. 6. Schematic of system.

Fig. 4(*b*), taking into account the various circuit tolerances and allowing a reasonable safety margin at the peak and valley current limits. The detailed analysis is given in Appendix 2. This shows that the optimum bias resistor for a given peak current, I_{pk}, is inversely proportional to I_{pk}, and that the maximum permissible read current is proportional to I_{pk}.

The word bias resistor, R_b, is then selected to ensure that the nominal bias voltage is such as to allow the use of the optimum bias resistor when the final choice of this resistor is determined by the availability of standard components. The minimum values of the operating currents, I_S and I_C, are then determined.

The results obtained in Appendix 2 which are relevant to a peak current of 5 mA and a word length of 24 digits are summarized below.

Peak current storage diodes	5 mA
Bias resistor	2 kΩ
Maximum read current	2·1 mA
Safety factor, Δi_S	0·5 mA
Word bias resistor	290 Ω
Clear current (I_C)	116 mA
Select current (I_S)	175 mA

Note:

(i) Δi_S is the factor of safety obtained at the peak and valley current limits under the 'worst case' operating conditions, and allows satisfactory operation in an ambient temperature range of 25–40 deg C.

(ii) The calculations of Appendix 2 are based on the following figures:

bias voltage $V_b = 3{\cdot}82 \text{ V} \pm 10\%$

$$\left(\frac{I_{pk}}{I_v}\right)_{\min} = 5$$

tolerance on $I_{pk} = \pm 5\%$

(iii) A tolerance of $\pm$ 3% is allowed on all resistors to allow for drift in the value of the 1% resistors used in practice. The maximum negative supply voltage, $-E$ in Fig. 5(*a*), is -24 V $\pm$ 0·5 V.

4.2. *Word Selection Mechanism*

The word selection mechanism employs a transistor matrix assembly which is preferred on the grounds of speed and simplicity to the less expensive but slower diode and diode-transformer systems also investigated. The mechanism is outlined in Fig. 7. The Y drivers are the voltage switches and the X drivers current switches for the matrices made up of the emitter-base

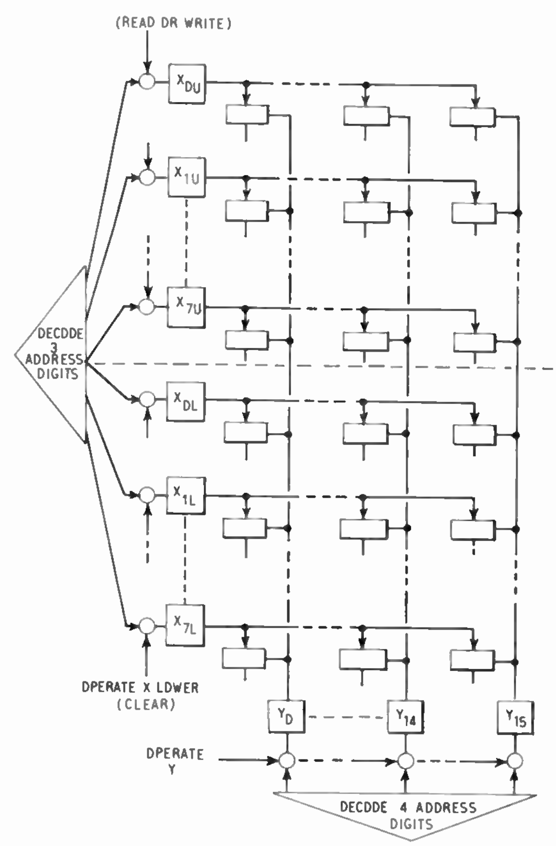


Fig. 7. Word selection mechanism.

diodes of the transistors which correspond to the switches SU and SL in Fig. 5(*a*). The collector of each transistor is connected to either an upper or lower word wire in the array of storage elements (Fig. 6), and the various store cycles are obtained by operating the upper or lower set of X drivers in the required sequence.

4.3. *Digit Circuitry*

The digit circuitry comprises the read and write switches, the read amplifier and OR gate for combining the outputs from the 4 sections of the digit line, the discriminator and digit staticizor. The design of this part of the system is considerably influenced by the choice of read diode, which has the following primary requirements:

(i) small variation between diodes in the forward voltage drop when carrying the read current, since this voltage is a component of the read out signal of the system.

(ii) low leakage current, which effectively subtracts from the read current and is voltage and temperature sensitive.

(iii) low reverse bias capacitance and

(iv) low hole storage characteristics.

This last factor is important because when the digit wire is returned to its quiescent potential after a read or write operation a reverse current flows in the selected read diode which can destroy the stored information. The system is most sensitive to the effect immediately after a 'write 1' operation and the difficulty is overcome by limiting the current available to return the digit wire to its quiescent level. The process then occurs over a significant period of time, and the hole storage charge decays relatively slowly with only a small transient reverse current. It is necessary to select a read diode with a low hole storage factor if adequate operating margins are to be achieved. For this reason the AAZ13 diode is preferred to a diode of the OA47 type which has better leakage current and reverse-bias capacitance characteristics. The forward voltage drop requirement can be met with this diode as the manufacturers can supply units with a tolerance as low as $\pm$ 30 mV at low currents. Some details of the digit circuits are given in Appendix 3.

4.4. *Physical Construction*

The storage elements are mounted in a three dimensional array in which the digit wire is backed by an earth plane. The length of each of the four sections of a digit wire is some twelve inches and this contributes about one third of the total capacitance of a digit wire section. In the experimental arrangement the complete storage array of 128 twenty-four digit words is contained in a volume of approximately 14 in × 24 in × 8 in (35 cm × 61 cm × 20 cm).

5. Practical Results

Much of the preliminary work on the system has been concerned with establishing the fundamental operating margins in terms of the currents involved. The results of these investigations are now considered.

5.1. *Read Current Limits*

The upper limit of the read current amplitude is determined by the sensitivity of the low voltage storage state as the peak current is approached. Near the peak current it is found that the storage diodes are liable to switch with small random disturbances in the system, and considerable care must be taken to minimize these effects. The disturbances can arise by radiation effects from nearby electrical devices and also appear on the power supply inputs to the system via the mains supply. The problem is resolved by efficient filtering of the power rails and surrounding the storage array by a screened enclosure.

In the test used to establish this limit the worst possible condition of bias and peak current tolerance were simulated in a particular storage element, and the system was then cycled for a considerable time period. The results indicate that a read current of 2·5 mA is non-destructive under these conditions, although when the current is increased to 2·7 mA a failure rate of about 1 per hour is observed.

Two factors influence the minimum acceptable read current: Firstly, there is a reduction in the rate at which the digit levels are established at the input to the read amplifier, which can result in unsatisfactory performance because the output gate timing is set for the nominal read current. Secondly, there is a reduction in absolute output signal levels as the read current is reduced as the voltage drop across the storage and read diodes is a function of this current. It is difficult to separate the two effects although the first is probably more important. In practice, satisfactory operation is obtained with the read current as low as 1·5 mA.

5.2. *Reverse Digit Wire Current*

The current which returns the digit wire to its quiescent level after a read or write operation must be limited to minimize the hole storage current in the read diode. The maximum permissible amplitude of the current switched into the digit wire is a complex function of the digit line capacitance and the hole storage characteristics of the read diodes. No firm upper limit has been established for this current, but it is certainly greater than 6 mA with the AAZ13 read diodes arranged in groups of 32 as here. With OA47 type read diodes it was found impossible to store '1' digits when the reverse digit wire current exceeded 2 mA.

Table 2

Operating margins

Parameter	Nominal value (mA)	Limits (mA) Upper	Lower	Conditions	Comments
Read current	2	2·5		Maximum bias current and minimum peak current.	Repeated access of stored '0' digit satisfactory over long period.
			1·5	Output gate timing set for nominal read current. Nominal elsewhere.	Output of read amplifier becoming unsatisfactory.
Reverse digit line current	2	>6		Nominal elsewhere. AAZ13 read diodes.	Reading and writing '1' digits. No firm upper limit.
Clear current I_C	125		100	Nominal elsewhere.	Cyclic test reversing '1's and '0's through a number of words. Occasional poor clear operation.
Selection current I_S	200	250		Nominal elsewhere.	Upper word wire tolerance limits exceeded.
			160		Occasional poor selection waveform.

5.3. *Operating Currents*

A number of cyclic tests have been made to establish the margins on the clear current, I_C, and the selection current, I_S. The nominal values of these currents are made rather greater than the theoretical minimum values obtained in Appendix 2, and the system is found to be relatively insensitive to variations in these currents.

The results of the experiments described in the preceding Section are summarized in Table 2.

5.4. *Operating Speed*

Tests have been made on a simulated full scale system to establish the minimum access and cycle times which are consistant with reliable operation. In a read order an access time of 150 mμs is realized with a cycle time of 250 mμs. The write cycle time is about 330 mμs. Photographs showing the waveforms obtained in various parts of the system are given in Fig. 8.

6. Conclusions

A particular design for a tunnel diode store with non-destructive read-out has been described and its feasibility investigated. The investigations demonstrate that such a system is one solution to the problem of providing the fairly small capacity high speed stores required in the modern computing machine.

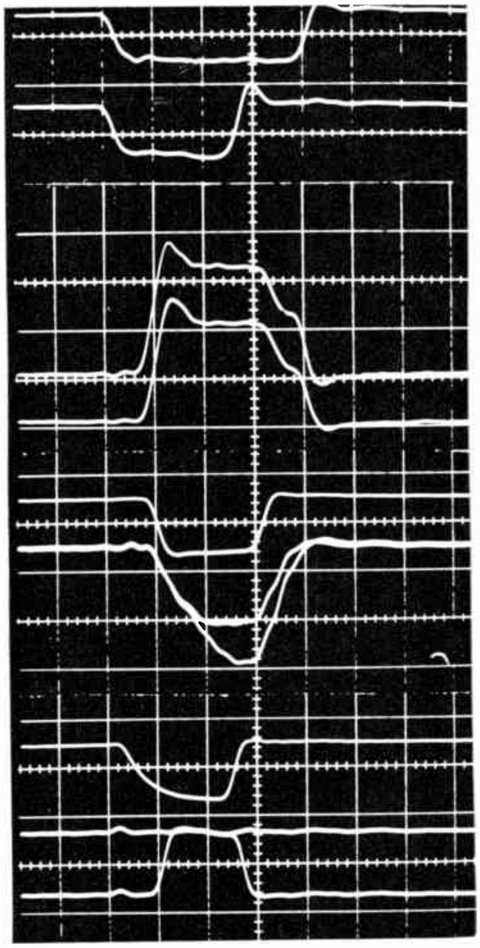
In the context of the *Atlas* requirements the demanded access and cycle times can be obtained in a unit of 128 words of 24 digits per word, since the read cycle time is 0·25 μs and the corresponding figure for the write cycle time is 0·33 μs (see Fig. 8). The advantage of the non-destructive read-out feature is not as significant as was at first expected because the write cycle, which is not necessarily limited by stray capacitance, is only about 1·25 times longer than the read cycle.

The techniques employed are quite costly, and in particular the word selection mechanism requires approximately 300 expensive transistors. However, the digit circuitry is relatively simple and inexpensive. This indicates that the system is not an economic proposition for storage capacities greatly in excess of 128 words.

An assessment has been made of the potentialities of this system in terms of a much smaller store of capacity 16 words of 25 digits. Here the word selection currents are unchanged but the word selection matrix can be eliminated and some time gained in exchange for a slightly increased cost per word. The digit wire capacitance is of course much reduced. With this arrangement a read cycle time of some 125 mμs is indicated and the write cycle is about 1·5 times longer than this.

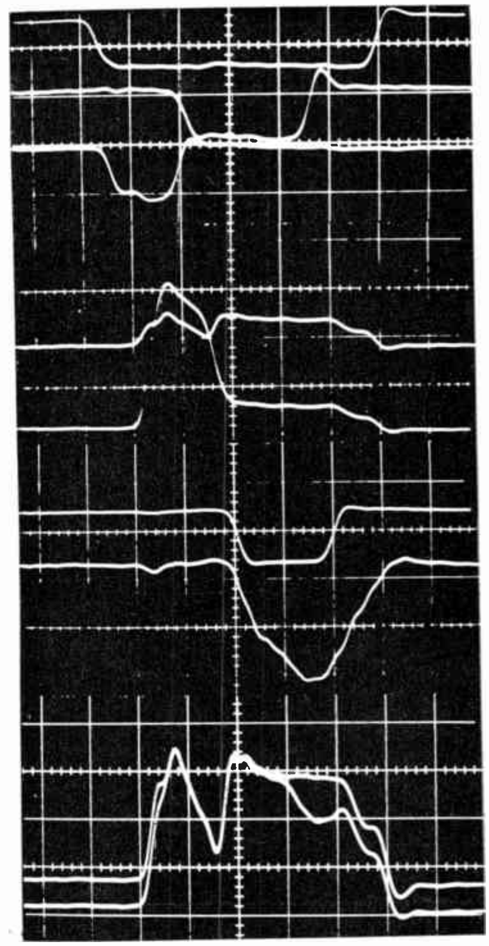
7. Acknowledgments

The authors wish to acknowledge the encouragement and facilities provided by Professor T. Kilburn, and the assistance of Messrs. J. Roach and J. B. Earnshaw who were responsible for most of the practical work described in this paper.



(a) Read cycle

(i) Operate Y
(ii) Operate X (upper)
(iii) Upper word wire (0·5 V/division)
(iv) Lower word wire (0·5 V/division)
(v) Digit strobe
(vi) Output gate
(vii) Staticizor output for '0' and '1'



(b) Write cycle

(i) Operate Y
(ii) Operate X (upper)
(iii) Operate X (lower)
(iv) Upper word wire (2 V/division)
(v) Lower word wire (2 V/division)
(vi) Digit strobe
(vii) Digit wire
(viii) Anode of storage diode (0·5 V/division)
(ix) Cathode of storage diode (0·5 V/division)

Fig. 8. Waveforms. (Time scale = 50 mμs/division).

8. References

1. D. B. G. Edwards, M. J. Lanigan and T. Kilburn, "Ferrite-core memory systems with rapid cycle times", *Proc. Instn Elect. Engrs*, **107**, Part B, No. 36, pp. 585–98, November 1960.
2. L. Esaki, "New phenomenon in narrow germanium p-n junctions", *Phys. Rev.*, **109**, No. 2, pp. 603–4, 15th January 1958.
3. T. Kilburn, D. B. G. Edwards, M. J. Lanigan, F. H. Sumner, "One level storage system", *Trans. Inst. Radio Engrs (Electronic Computers)*, EC–**11**, No. 2, pp. 223–35, April 1962.
4. M. J. Lanigan, "The Organization of the Main Core System in the *Atlas* Computer". Ph.D. Thesis (Part II), Manchester University, 1962.
5. G. R. Hoffman, J. A. Turner and T. Kilburn, "High speed digital storage using cylindrical magnetic films", *J.Brit.I.R.E.*, **20**, No. 1, pp. 31–6, January 1960.
6. Rajchman, J. A., "Computer memories; a survey of the state of the art", *Proc. Inst. Radio Engrs*, **49**, No. 1, pp. 104–27, January 1961.
7. G. B. B. Chaplin and P. M. Thompson, "A fast word-organized tunnel diode memory using voltage mode selection". Digest of Technical Papers, International Solid State Circuits Conference, February, 1961.
8. S. Takahashi, O. Ishii, *et al.*, "Tunnel diode high speed memory". Electrotechnical Laboratory, Tokyo, Japan, 1961.

9. M. J. Lanigan, "A Tunnel Diode Storage System with Non-Destructive Read-Out". *Op. cit*, (Part III).

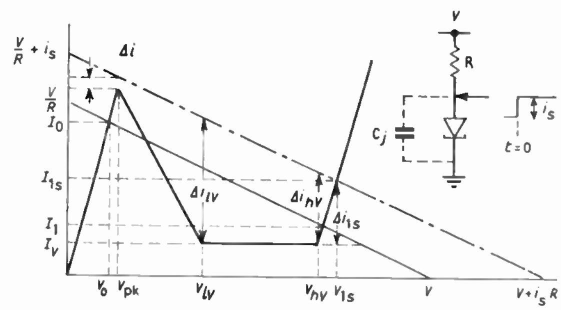
10. M. J. Lanigan and D. B. G. Edwards, "Digital information storage apparatus". Provisional Patent Application No. 6633/62. (February 1962).

9. Appendix 1

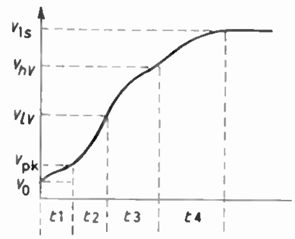
An Approximate Analysis for the Switching Time from Low to High Voltage

The problem considered here is that of tunnel diode biased in the low-voltage state, which is switched to the high voltage state by the action of a step function of current i_s. The analysis assumes a constant junction capacitance C_j and a linear tunnel diode characteristic, Fig. 9. The switching interval, from the low voltage state V_0 to the high voltage state, V_{1s}, may be considered in four separate regimes corresponding to the four linear divisions of the characteristic. The switching waveforms, Fig. 9(*b*), follows an exponential path in each regime in accordance with the equations below:

$$V_0 \to V_{pk}; \qquad v(t) = V_0 + \frac{i_s R r_1}{R + r_1}(1 - e^{-t/T_1}) \qquad \ldots\ldots(1)$$



(*a*) Approximate characteristic.



(*b*) Switching waveform.

Fig. 9. Tunnel diode switching

Table 3

Switching times

$I_0 > I_1$	$I_0 \simeq I_1$ (Constant current bias)	Notation
$t_1 = T_1 \ln\left[\dfrac{1}{1 - \dfrac{I_{pk}-I_0}{i_s}\dfrac{R+r_1}{R}}\right]$	$t_1 \simeq \dfrac{q_1}{I_{pk}-I_0}\ln\left[\dfrac{i_s}{\Delta i}\right]$	$r_1 = \dfrac{V_{pk}-V_0}{I_{pk}-I_0}$ $T_1 = C_j\dfrac{Rr_1}{R+r_1}$ $q_1 = C_j(V_{pk}-V_0)$
$t_2 = T_2 \ln\left[1 + \dfrac{(I_{pk}-I_v)(R-\bar{r})}{\Delta i R}\right]$	$t_2 \simeq \dfrac{q_2}{I_{pk}-I_v}\ln\left[\dfrac{\Delta i + I_{pk} - I_v}{\Delta i}\right]$	$\bar{r} = \dfrac{V_{lv}-V_{pk}}{I_{pk}-I_v}$ $T_2 = C_j\dfrac{R\bar{r}}{R-\bar{r}}$ $q_2 = C_j(V_{lv}-V_{pk})$
$t_3 = T_3 \ln\left[\dfrac{\Delta i_{lv}}{\Delta i_{hv}}\right]$	$t_3 \simeq \dfrac{q_3}{\Delta i_{hv}}$	$T_3 = C_j R$ $q_3 = C_j(V_{hv}-V_{lv})$
$t_4 = T_4 \ln\left[\dfrac{1}{1 - \dfrac{\Delta i_{1s}}{i_{hv}}\dfrac{R+r_4}{R}}\right]$	$t_4 \simeq 2{\cdot}3\dfrac{q_4}{I_{1s}-I_v}$	$r_4 = \dfrac{V_{1s}-V_{hv}}{I_{1s}-I_v}$ $T_4 = C_j\dfrac{Rr_4}{R+r_4}$ $q_4 = C_j(V_{1s}-V_{hv})$

$$V_{pk} \to V_{lv}; \quad v(t) = V_{pk} + \frac{\Delta i R\bar{r}}{R-\bar{r}}(e^{t/T_2}-1) \quad \ldots\ldots(2)$$

$$V_{lv} \to V_{hv}; \quad v(t) = V_{lv} + \Delta i_{lv} R(1-e^{-t/T_3}) \quad \ldots\ldots(3)$$

$$V_{hv} \to V_{ls}; \quad v(t) = V_{hv} + \frac{\Delta i_{hv} R r_4}{R+r_4}(1-e^{-t/T_4}) \ldots\ldots(4)$$

The notation here is defined in the summary of results in Table 3. Equation (2) which is that of an increasing exponential, is of particular interest as it illustrates the effect of the negative conductance region between V_{pk} and V_{lv}. Here as the diode voltage increases an increasing amount of current is diverted into the junction capacitance and the switching rate increases with time Fig. 9(*a*). The equations (1) to (4) allow the transition times in each region to be formulated; these results are tabulated in Table 3.

In the system described in Sections 3.2 and 4 the 5 mA storage diodes are biased with a constant current of 2 mA and are written to the '1' state by the action of a 4 mA trigger current. Typical figures for the diodes are:

$V_{pk} = 50$ mV $\qquad V_{lv} = 240$ mV

$V_{hv} = 380$ mV $\qquad V_{ls} = 480$ mV (at 6 mA)

and $C_j = 30$ pF

Then setting

$V_0 = 20$ mV at $I_0 = 2$ mA

$I_v = 0{\cdot}8$ mA

yields

$q_1 = 0{\cdot}9$ pC $\qquad q_2 = 5{\cdot}7$ pC

$q_3 = 5{\cdot}2$ pC $\qquad q_4 = 3$ pC

and the corresponding switching times are

$t_1 = 0{\cdot}42$ mμs

$t_2 = 2{\cdot}25$ mμs

$t_3 = 1{\cdot}0$ mμs

$t_4 = 1{\cdot}3$ mμs

The typical total switching time observed when the triggering current is a step function is about 5 mμs, which compares favourably with the theoretical predictions. In the storage system the switching time is rather longer than 5 mμs (8 to 10 mμs) because of the extra stray capacitance across the diode and slower triggering current waveform.

10. Appendix 2

Storage Element Bias Resistor and Operating Currents

In Fig. 4(*b*) the position of the load line, defined by the bias voltage V_b and the resistor R, is determined by two factors:

(*a*) the quantity $(I_{pk}-I_0)$ should be as large as possible so that the non-destructive read current can be large and

(*b*) the stability of the two storage states of the element must be assured in the worst conditions of diode and circuit tolerance.

In practice the ideal solution of biasing with a constant current which is marginally greater than the maximum specified valley current cannot be achieved. A compromise is therefore adopted which allows for an approximate 10% difference in the diode currents between the two stables states.

10.1. *Choice of Bias Resistor and Nominal Bias Voltage*

The stability of the high voltage state is obtained by satisfying the inequality

$$\frac{V_{b(\min)} - V_{v(\max)}}{R_{\max}} \geqslant \frac{I_{pk}}{\gamma_{\min}} + k_s I_{pk} \quad \ldots\ldots(5)$$

where γ is the ratio I_{pk}/I_v which is expressed in terms of the nominal peak current, and k_s is a factor of safety.

Let the tolerance on V_b be $\pm\, t_b V_b$

" " " " R be $\pm\, t_R R$

" " " " I_{pk} be $\pm\, t_{pk} I_{pk}$

then (5) may be rewritten

$$R \leqslant \frac{V_b(1-t_b) - V_{v(\max)}}{I_{pk}\left(\frac{1}{\gamma_{\min}} + k_s\right)(1+t_R)} \quad \ldots\ldots(6)$$

Now the quantity $(I_{pk}-I_0)$ has its maximum value when R has the maximum value allowed by eqn. (6) i.e. the optimum nominal value of the bias resistor is given by

$$R = \frac{V_b(1-t_b) - V_{v(\max)}}{I_{pk}(\gamma_{\min}+k_s)(1+t_R)} \quad \ldots\ldots(7)$$

The final choice of R is determined by the availability of standard components, and the values of R tabulated in Table 4 satisfy eqn. (7) with the following parameters:

$\gamma_{\min} = 5$ $\qquad$ from the manufacturer's data.

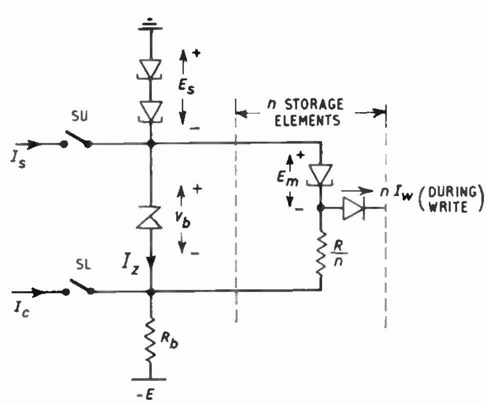


Fig. 10. Storage element bias resistor and operating currents.

$V_{v(\max)} = 0.35$ V from the manufacturer's data.

$k_s = 0.1$

$t_R = 0.03$ this value allows for possible drift in the value of the 1% resistors used in practice.

$t_b = 0.1$ this tolerance is assumed at this stage.

$V_b = 3.82$ V

The required value of V_b is independent of I_{pk} because $R.I_{pk}$ is constant. The value of V_b demands a nominal Zener diode current of 24 mA in KS30A diodes used. This is obtained by choosing an appropriate value for the resistor R_b (Fig. 10).

10.2. *Choice of Word Bias Resistor*

In Fig. 10 the Zener current, I_Z, is given by

$$I_Z = \frac{E-(V_b+E_s)}{R_b} - \frac{(V_b-E_m)}{R/n} \qquad \ldots\ldots(8)$$

The resistor R_b is chosen so that the nominal Zener current and bias voltage are obtained when:

(*a*) E and R have their nominal values,

(*b*) E_m has a value which represents an equal distribution of high and low voltage storage states in the n storage elements, and

(*c*) the word is midway between the unselected and fully selected states.

Under these circumstances

$E = 24$ V

$R = \dfrac{10}{I_{pk}}$ ohms

$E_m = 0.22$ V

$E_s = 0.44$ V

and $\left.\begin{array}{l} I_Z = 24 \text{ mA} \\ V_b = 3.82 \text{ V} \end{array}\right\}$ by definition.

The corresponding values of R_b are given in Table 4.

10.3. *Variation of Bias Voltage*

In the calculations of Section 10.1 it was assumed that the bias voltage could vary ± 10% about its nominal value. The actual variation which can occur under operating conditions is estimated by substituting a linear approximation for V_b in terms of I_Z in eqn. (8). This is

$$V_b = E_Z + I_Z r_Z \qquad \ldots\ldots(9)$$

where $r_Z = 16$ ohms and $E_Z = 3.45$ V.

Equation (9) represents the avalanche characteristic of a typical KS30A diode to an accuracy of better than ±2% for current variations of 10 mA on either side of the nominal Zener current of 24 mA. Equation (8) then becomes

$$I_Z = \frac{(E-E_s-E_Z)R - nR_b(E_Z-E_m)}{RR_b + r_Z(R+nR_b)} \qquad \ldots\ldots(10)$$

Setting $E = 24 \text{ V} \pm 0.5$ V

$E_s = 0.88 \text{ V} \pm 0.1$ V

$E_Z = 3.45 \text{ V} \pm 5\%$ V (manufacturer's tolerance)

$E_m = 20$ mV or $0.44 \text{ V} \pm 30$ mV

Allowing a ± 3% tolerance on all resistor values, enables the minimum and maximum values of I_Z to be obtained. These are tabulated in Table 4 and the corresponding variation in V_b is also shown. In all cases for $I_{pk} \leqslant 5$ mA, this is within ± 10% of nominal, which indicates that the allowed tolerance on V_b used in the preceding calculations is justified.

10.4. *Operating Currents*

10.4.1. Read current

The maximum read current is taken as that which gives the same factor of safety at the peak current limit as exists at the valley current, i.e.

$$i_{s(\max)} = I_{pk(\min)} - \frac{V_{b(\max)} - V_{0(\min)}}{R_{\min}} - k_s I_{pk} \qquad \ldots\ldots(11)$$

The values of maximum read current tabulated in Table 4 are obtained from eqn. (11) by inserting the appropriate values and setting $k_s = 0.1$.

Table 4

Design values for various peak currents

Peak current (mA)	Nominal bias resistor R kΩ	Nominal bias voltage V_b	Nominal Zener current I_s mA	Nominal word bias resistor R_b ohms	Variation of I_Z (mA) max.	Variation of I_Z (mA) min.	Variation of V_b (volts) max.	Variation of V_b (volts) min.	% Variation of V_b	Max. read current mA	Min. clear current I_C mA	Min. selection current I_s mA
5	2	3·82	24	290	31	10·5	3·93	3·5	+3 −8·5	2·1	116	175
3	3·33	3·82	24	390	29	22·8	3·88	3·75	+1·6 −1·9	1·25	90	115
1	10	3·82	24	600	26·7	20	3·85	3·75	+1·9 −1·9	0·42	65	56

10.4.2. Clear current

The current I_C switched into the lower word wire (Fig. 10) when the 'clear to zero' operation is initiated has three components. These provide (*a*) the current flowing in the resistor R_b, (*b*) a small reverse current in each storage element, and (*c*) a current to reverse the bias voltage developed across the Zener diode in an acceptable time. The magnitude of this last component is governed by a 'charge constant', ΔQ_Z, which relates the switching time of the Zener diode to the switching current. The value of ΔQ_Z varies widely between Zener diodes of the same nominal voltage from different manufacturers. For the KS30A diode ΔQ_Z is some 400 pC, and switching times of about 20 mμs are obtained with a 20 mA switching current.

The value of I_C must satisfy the relationship

$$I_C \geqslant \frac{v_L + E}{R_b} + \frac{v_L}{R/n} + \frac{\Delta Q_Z}{t_s} \qquad \ldots\ldots(12)$$

where t_s is the required switching time, and v_L is the voltage level of the lower word wire. The final value of v_L is chosen to be $+0{\cdot}5$ V, since the forward current in the Zener diode is then very small, (of the order of 1 mA) and hole storage effects when I_C is removed may be neglected. Then setting

$$v_L = +0{\cdot}5 \text{ V}$$
$$\Delta Q_Z = 400 \text{ pC}$$
$$t_s = 20 \text{ m}\mu\text{s}$$

in eqn. (12) yields the minimum values of I_C tabulated in Table 4.

10.4.3. Selection current

The selection current I_s switched into the upper word wire (Fig. 10) must be large enough to ensure that the selection tunnel diodes conduct in the reverse direction when all the digits of the word are being written to high voltage state. That is

$$I_s \geqslant \frac{E - V_b}{R_b} + nI_W \qquad \ldots\ldots(13)$$

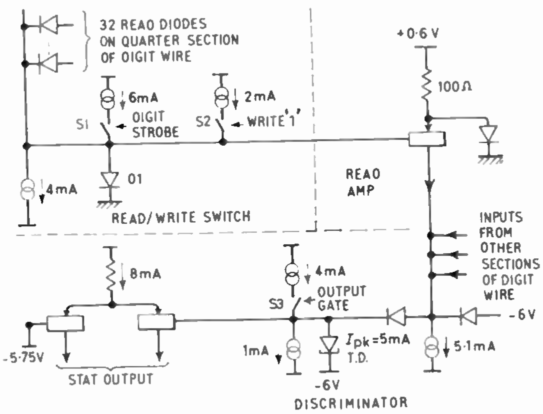
where I_W is the digit write current, which is made equal to $0{\cdot}8I_{pk}$. The minimum values of I_s given in Table 4 are then obtained by inserting the most unfavourable tolerance conditions in eqn. (13).

11. Appendix 3

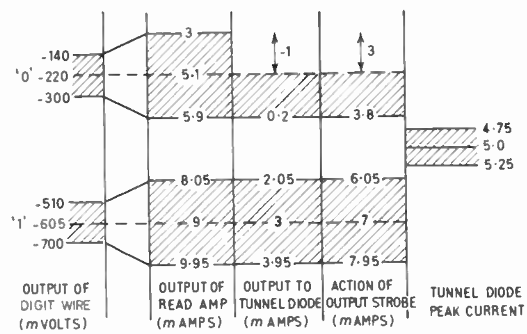
Digit Circuits

Each of the four sections of a digit wire is associated with a read/write switch, which is operated by a version of the digit strobe assigned to the appropriate page position (see Fig. 6), and a read amplifier. The outputs of the four read amplifiers for a particular digit wire feed a common discriminating and output staticizor circuit. A schematic of the system is given in Fig. 11(*a*).

Under quiescent conditions switches S1, S2 and S3 are closed and the input to the read amplifier is held at about $+0{\cdot}6$ V by conduction of the silicon diode D1. At the start of each store cycle S3 is opened by the action of the output gate and the tunnel diode, TD, is switched to a low voltage state by a 1 mA reset current.



(*a*) Schematic for one quarter digit wire.



(*b*) Effect of output tolerances.

Fig. 11. Digit circuits.

11.1. *Read Cycle*

In a read cycle switch S2 remains closed and the appropriate digit strobe is operated as the addressed upper word wire reaches the selection level. This opens S1 and a read current of 2 mA is extracted from a section of the digit wire. The input level of this read amplifier then becomes more negative until the selected read diode is carrying the read current. The read amplifier input is subject to tolerance variations which are summarized in Table 5.

Table 5

Digit level tolerances

	Potential (mV) '0' Digit	'1' Digit
Upper word wire	+ 115 ± 35	+ 115 ± 35
Storage diode at 4 mA	− 45 ± 15	− 430 ± 30
Read diode at 2 mA	− 290 ± 30	− 290 ± 30
Read amplifier input	− 220 ± 80	− 605 ± 95

At a sufficiently negative input level the read amplifier has a well defined transfer characteristic of 10 mA/V and the action of the amplifier and discriminating circuit under the various conditions of input tolerance is illustrated in Fig. 11(*b*). A decision is obtained from the discriminating circuit when the output gate is operated to close S3. The timing of the output gate (positive-going edge) is shown in Fig. 8. The information read out of the store is retained at the output of the staticizor by the storage action of the discriminating tunnel diode until S3 is re-opened at the beginning of the next store cycle.

There is an adequate safety margin in the discriminating system to cope with minor inaccuracies of definition of the currents which control its action. A degree of control in the discrimination level is possible by variation of the 1 mA reset current and this facility is incorporated in the system to establish operating margins.

11.2 *Write Cycle*

In a write cycle the information to be written into the store controls the action of the switch S2, which is opened in those digit positions where writing to a '1' is required. In this way operation of the digit strobe causes 4 mA to flow from the digit wire and the selected storage diodes are set to a high voltage condition.

Manuscript received by the Institution on 22nd April 1963. (Paper No. 859/C60).

STANDARD FREQUENCY TRANSMISSIONS

(*Communication from the National Physical Laboratory*)

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

October 1963

1963 October	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	1963 October	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	− 130·3	− 129·3	− 16	17	− 130·8	− 130·6	− 8
2	− 130·0	− 130·7	− 15	18	− 130·5	− 131·0	− 4
3	− 129·5	− 130·9	− 14	19	− 130·5	− 130·1	− 4
4	− 130·6	− 130·0	− 15	20	− 130·4	− 130·9	− 4
5	− 130·3	− 130·3	− 13	21	− 130·2	− 130·9	− 6
6	− 129·9	− 130·2	− 12	22	− 130·6	− 130·5	− 6
7	− 129·9	− 131·0	− 13	23	− 130·4	− 130·9	− 4
8	− 130·1	− 130·4	− 11	24	− 130·3	− 130·7	− 4
9	− 129·9	− 129·6	− 11	25	− 130·2	− 130·5	− 3
10	− 129·9	− 130·1	− 11	26	− 130·2	− 129·9	− 1
11	− 130·4	− 130·4	− 9	27	− 130·7	− 129·9	—
12	—	− 130·4	− 9	28	− 130·5	− 130·6	0
13	− 130·4	—	—	29	− 129·9	− 130·6	0
14	—	− 129·5	− 9	30	− 130·0	− 130·8	0
15	− 129·8	− 130·3	− 8	31	− 131·5	− 130·9	0
16	− 130·7	− 130·3	− 6				

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

Note: the phase of the GBR/MSF time signals was retarded by 100 milliseconds at 00 00 UT on 1st November, 1963.

U.D.C. 538.54: 620.1

The Development of Eddy-current Testing Techniques for Tube Inspection

By
D. TERRY†

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

Summary: The paper outlines the need for a high speed method of examining tubes for defects and discusses to what extent eddy-current techniques meet this need. A short history is given of the introduction of eddy current testing used for the examination of solid and hollow cylinders. Eddy current testing principles are explained, together with methods of measuring the effects of variations in a test specimen, with the aid of block circuit diagrams and complex impedance curves.

The merits of a.c. and d.c. saturation techniques are dealt with and practical testing equipment which employs phase and modulation analysis is described. The application of phase sensitive devices and modulation analysis techniques is explained and the limitations assessed. The uses of internal probes and surface probes are briefly described.

1. Introduction

The safety factor of a component can only be reduced if strict measures are taken to guarantee that the material has consistent physical properties and is free from harmful defects. These factors have been controlled in the past by selecting representative samples from the bulk production and subjecting them to as many different destructive tests as possible. The obvious snag with this system is the possibility of a sub-standard component passing through which does not meet essential requirements and fails in service. With domestic appliances such as a vacuum cleaner or refrigerator, no real harm is done, but an aircraft fitted with a faulty turbine blade or fuel pipe can bring disaster. The effort needed to develop inspection techniques for tubes in aircraft can be put to good use in other fields of industry.

Non-destructive testing is a combination of science and art; it is a subject found to exasperate many of those who have dealings with it, because the conditions found in practical application do not measure up in every respect to the scientific concept on which a particular method is based. Eddy-current testing can be considered as being particularly irksome in this respect.

The basic principles of eddy-current testing have been stated many times but practical techniques have been very slow to follow. For instance the efficient examination of tubes manufactured in ferro-magnetic steel or stainless steel has only been possible in recent years. The reluctance to adopt eddy-current testing as a means of examination can be attributed to several factors. In the first instance incorrect conclusions were made from very early trials before a real understanding of the theory was appreciated. Many manufacturers of equipment and users were over optimistic and made extravagant claims; others were over cautious in applying the method and did not make any real contributions. Modern practices for manufacturing electronic components has led to reliable testing equipment requiring little maintenance and service. Improved circuits and new devices have given a new slant to the subject, allowing the user to obtain more information about the test.

The largest single contribution in this field has been made by F. Förster and his collaborators who have produced a very wide range of practical testing equipment of robust design suitable for works use.[1, 2, 3] P. Graneau, S. A. Swan and R. Hochschild have also contributed in helping to develop the theoretical basis for eddy-current analysis which support the experimental results obtained by Förster.[4, 5, 6, 7] Eddy-current testing is generally associated with the detection of defects in metals, but other applications include thickness measurement of metals on non-metals and vice-versa, determination of electrical conductivity and the detection of tramp metal.[8]

2. Principles of Testing

When a cylinder of metal is placed inside a coil fed with alternating current, the resistive and reactive components of the coil will appear to be modified. The resistance and reactance of the metal cylinder, which may be considered as the secondary winding of a transformer having one turn, can be measured by the surrounding coil in much the same way as the

† Accles & Pollock Ltd., Oldbury, Birmingham.

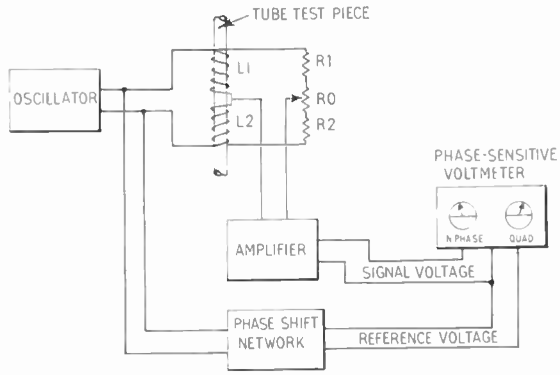


Fig. 1. Block diagram of equipment for measuring the effects of variations in dimensions, conductivity and permeability of a tubular test piece.

equivalent primary impedance is determined for a transformer on a short-circuit test. The change in coil impedance between the empty and loaded condition depends upon the dimensions, electrical conductivity and permeability of the cylinder, the coil shape and the degree of electromagnetic coupling between the coil and cylinder (usually called the 'coil filling factor'). The presence of a defect in the cylinder also causes a corresponding change in coil impedance.

The extent of changes in coil impedance through variations in cylinder dimensions, conductivity, permeability or the presence of defects, can be measured on equipment similar to that shown by the block circuit diagram in Fig. 1. The output from a variable oscillator feeds two test coils of inductance L_1 and L_2 connected in a bridge circuit with resistances R_1 and R_2 as the other two arms and resistance R_0 adjusted to give a balanced condition. Two methods of measurement are possible; the first with L_1 and L_2 arranged as shown in the diagram (known as the auto-comparison method), the second with L_1 and L_2 separated (known as the comparison method). When the auto-comparison method is used, the characteristics of the sample inside L_1 are balanced against those of the adjacent sample in L_2. The comparison method requires a standard sample inside L_1 with separate samples inserted in L_2, possessing the characteristics to be studied. Specimens must be selected which differ from the standard in one respect only and a great deal of additional measurements are needed to ensure this. For instance when a study is being made of the effects of dimensional changes, strict control must be kept on heat treatment or cold work to maintain constant conductivity and permeability. Variations in the specimen result in a bridge output which is amplified and fed to a phase sensitive voltmeter or cathode ray tube, together with a reference voltage from the oscillator via a phase sensitive network.

Controlled experiments and theory have shown that the impedance of a test coil surrounding a cylinder, whose magnetic permeability is unity, varies in accordance with the curves shown in Fig. 2. For convenience the reactive and resistive components of impedance for the loaded condition have been normalized to the reactive impedance of the coil in the absence of the metal cylinder. Curves are shown for the case where the cylinder completely fills the coil. The abscissa values are in terms of $(R-R_0)/\omega L_0$ and the ordinate values in terms of $\omega L/\omega L_0$, where R and ωL are the resistive and reactive components for the coil with the specimen inserted, R_0 and ωL_0 is the resistance and reactance of the coil with the specimen removed.

Most of the practical work concerned with impedance analysis has been carried out at the Institut Dr. Förster, Reutlingen, Germany. This Institut was responsible for simplifying the application of the theory by introducing expressions which they called the 'limiting frequency' or the 'critical frequency' for

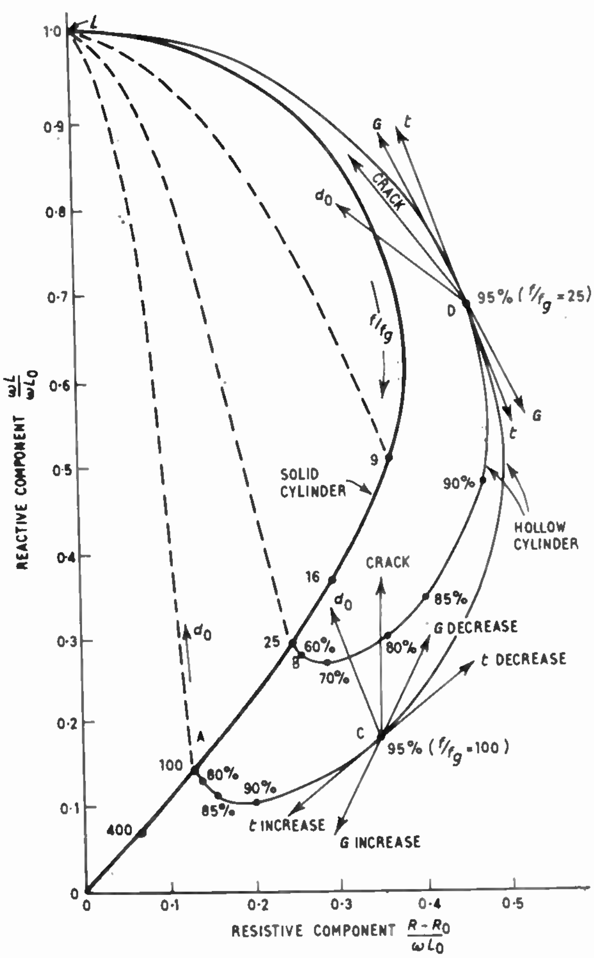


Fig. 2. Impedance plane for a coil encircling tubes when the inside diameter increases from zero in % d_0 for f/f_g ratios for 100 and 25. The effect of d_0, G and t variation is shown for the case of a thin walled tube.

various uniform shapes such as plate, sheet, spheres, solid and hollow cylinders. Theoretical work agreed very closely with results found by experiment, consequently the expressions developed are very convenient to use when determining the ideal test frequency for a uniform shaped object.[1]

2.1. *Solid Cylinders*

The 'limiting frequency' for a solid cylinder may be found from the expression:

$$f_g = \frac{5066}{G.\,\mu.\,d_0^2} \qquad \ldots.(1)$$

where f_g = limiting frequency in c/s

G = electrical conductivity in metre/ohm-mm²

d_0 = outside diameter of the cylinder in cm.

μ = relative magnetic permeability.

The 'limiting frequency' is characterized by the size and properties of the specimen under test and must not be confused with the testing frequency. The curves in Fig. 2 can best be explained by considering an example.

A solid cylinder of stainless steel, diameter 6 cm, conductivity 1·4 metre/ohm-mm² with a relative magnetic permeability of unity has an f_g value of 100 (calculated from equation (1)). At a test frequency f of 10 000 c/s, the ratio f/f_g would be 100, and represented on the impedance plane by point A, i.e. an empty coil is represented by point L and with the specimen in position by point A. Now if the test frequency were reduced to 2500 c/s, the new condition would be represented by point B where $f/f_g = 25$. Any reduction in outside diameter d_0, keeping G and μ constant causes point A to move along the dotted line towards point L (Fig. 2), with a phase angle difference of approximately 45 deg from the direction due to variations in conductivity.

By using phase-sensitive devices, conductivity variations can be separated from variations in outside diameter. This fact is made use of in some instruments designed for sorting mixed batches of metal components by virtue of their differences in electrical conductivity.

2.2. *Hollow Cylinders*

For a tube, the impedance plane is somewhat different from the impedance plane for a solid cylinder. Let us consider the previous example for a solid stainless steel cylinder with a diameter of 6 cm surrounded by a coil fed at a frequency of 10 kc/s. The normalized impedance is represented by point A on the impedance plane, Fig. 2, where $f/f_g = 100$. If the centre of the bar is now machined away until the inside diameter is 5·7 cm (95% d_0), the impedance will change to position C. Here vectors have been drawn to indicate the direction in which point C would move for changes in tube thickness, conductivity, outside diameter and the effect of a crack. Only one parameter can be considered at a time, keeping all the others constant. The direction shown for the crack effect can vary depending upon its depth, width and position in the tube wall. To detect defects in tubing satisfactorily by phase analysis methods, adequate separation between all the parameters is required. Even under these ideal conditions, when a 'measuring phase' is selected which is perpendicular say to the d_0 direction, the influences due to variations in conductivity and thickness are measured at the same time as influences due to defects. In practice, all these variations occur, possibly at the same time in some cases, making the detection of small defects impossible. At a lower testing frequency of 2·5 kc/s, $f/f_g = 25$, the same tube as before would take up position D on the impedance plane. Here the phase angle between all the parameters is smaller and becomes more difficult to separate than before. An appreciation of the methods used to manufacture tubing is important before deciding to apply phase analysis techniques, because a set of conditions for one batch may not be suitable for another batch of the same size and quality but made by a different process. One method of manufacture may produce a tube of constant internal diameter with slight variations in outside diameter, while another method may produce a tube of constant outside diameter with variations in internal diameter and yet still meet specification requirements.

2.3. *Impedance Analysis by the Ellipse Method*

When the signal output from the bridge in Fig. 1 is amplified and applied to the vertical plates of a cathode-ray tube and the reference voltage is applied to the horizontal plates, a stationary Lissajous pattern in the form of an ellipse will be produced.

The appearance of the ellipse depends on the amplitude and phase relation between the two voltages. Let us again consider the previous example of a stainless steel tube, outside diameter 6 cm, inside diameter 95% d_0, represented on the impedance plane by point C on Fig. 2. Voltage outputs from the bridge will have an amplitude and direction proportional to the impedance changes, and may be represented on a 'complex voltage plane' as shown in Fig. 3(*a*). Here a reference voltage OD has been selected which is in phase with the signal voltage caused by a reduction in d_0 and is applied to the XX plates of a c.r.t. in Fig. 3(*b*). Now when a signal caused by a reduction in d_0 is applied to the YY plates a tilted line will appear on the c.r.t. A crack giving a signal of the same amplitude as the reduction in d_0 but having a phase difference of ϕ_1 will produce an ellipse. A thickness reduction giving a similar signal

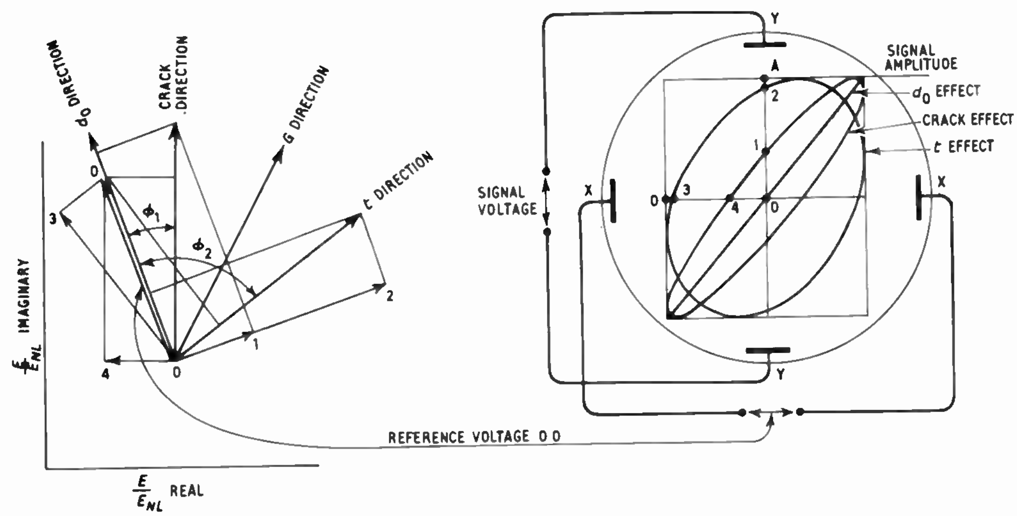


Fig. 3. Construction of the ellipse from a reference voltage and signal voltage for a thin walled tube.

amplitude, keeping d_0 constant, will produce an ellipse with a different shape to the one given by the crack.

When the signal and reference voltages are fed to a phase-conscious rectifier, the component of signal voltage which is perpendicular to the reference voltage becomes most important. With the particular case that has been chosen, the component due to a crack effect which is perpendicular to the d_0 direction is vector component O1 (OA sin ϕ_1), and the perpendicular component to the d_0 direction for thickness change is vector component O2, (OA sin ϕ_2). These components of the signal voltage are normally applied to a circuit which will trigger when a predetermined level has been reached, or alternatively fed to a pen recorder. The other points of interest on the ellipse are distance O4 which is equivalent to OD sin ϕ_1, and O3, equivalent to OD sin ϕ_2.

2.4. *Eddy-current Distribution*

Eddy currents induced by a coil surrounding a cylinder are not uniformly distributed over the entire cross-section but the eddy-current density is greatest on the outside surface and falls in its intensity with depth below the surface. This is commonly known as the 'skin effect'. The depth at which the eddy currents reach a value equal to 1/e times the surface value is one way of defining the depth of penetration.

For a semi-infinite plane conductor, the depth of penetration may be calculated from the following formula:

$$\text{depth} = \sqrt{\frac{\rho \times 10^3}{4\pi^2 \mu f}} \text{ cm}$$

where ρ = resistivity in microhm-cm
μ = relative permeability
f = testing frequency in c/s.

Table 1

Depth of penetration in inches for various materials

Material	Permeability	Resistivity	Frequency kc/s 0·05	0·5	1·0	10·0	100·0
Copper	1	1·65	0·357	0·113	0·080	0·025	0·008
Aluminium	1	2·88	0·474	0·150	0·106	0·033	0·010
Magnesium	1	4·35	0·582	0·184	0·130	0·041	0·013
Brass	1	7·5	0·755	0·240	0·169	0·054	0·017
Mild Steel	μ initial $\sim$ 500 can vary	20	0·055	0·017	0·012	0·004	0·0013
Mild Steel	at saturation $\Delta\mu \sim$ 1·0	20	1·230	0·380	0·268	0·090	0·029
Austenitic Stainless Steel	$\sim$ 1·0 can vary	73	2·380	0·755	0·532	0·168	0·053

Typical figures for the penetration depth of various materials and test frequencies are shown in Table 1. Förster has shown that this formula is only approximate for solid cylinders at ratios $f/f_g > 15$, but does not hold for lower ratios. Recent work by Bareham[9] has shown that the eddy-current distribution in thin tubes is vastly different to the values calculated using the above formula, and he shows that a much wider test frequency is possible than indicated by the 'skin-depth' equation. Since the signals given by defects in tubing are difficult to separate from the signals due to dimensional variations it is best to use the lowest test frequency that is permissible. Signals from bore defects would then have an amplitude comparable to signals from outside defects. Unfortunately as the frequency is lowered the electromagnetic coupling between coil and tube is poorer and defects become more difficult to detect. Therefore, a compromise must be reached on results obtained from practical investigation.

The problem of selecting the correct testing frequency for detecting longitudinal cracks and other defects in solid and hollow cylinders has been investigated in detail at the Förster Institut. Although the theoretical distribution of eddy currents have been determined, the effects of defects could not be accurately calculated. Förster decided to use a mercury cylinder as an ideal test piece and insert pieces of plastic of various sizes and shapes to represent defects. Measurements taken from a surrounding test coil gave the information needed, which could be interpolated for all materials of any diameter and thickness.[10, 11]

3. Testing Ferromagnetic Materials

When testing tubes manufactured in such materials as mild steel or stainless steel having ferromagnetic properties, the complex impedance plane previously considered for non-magnetic materials is changed considerably by the effect of magnetic permeability.[12] The inductance of a coil surrounding a cylinder is reduced when the material from which the cylinder is made is non-magnetic, such as copper, but increases for ferromagnetic materials.

The main problem in testing ferromagnetic material lies in the fact that the initial permeability depends upon chemical composition, metallurgical structure and most important of all on the presence of residual internal stresses introduced by the manufacturing process. These facts were discovered in the very early days of eddy-current testing and many unsuccessful attempts have been made to overcome the problem by impedance analysis. Alternative methods were resorted to, including techniques based on detecting the leakage flux from a defect while the specimen was situated in a magnetic field.[12, 17] Apart from the difficulty caused by permeability variations giving erroneous signals the depth of penetration of the eddy currents in ferromagnetic metals is very low when compared with other non-magnetic metals (see Table 1).

3.1. *A.C. Saturation*

By using a very high a.c. field strength the permeability in the outer layer of the tube surface is reduced to a low value every half cycle. The field strength decreases with depth from the surface due to the reaction of eddy currents in the outer layers of the tube. The permeability therefore increases with depth and eventually reaches a maximum value. Beyond the layer in the tube where this occurs, considerable resistance to deeper penetration is formed. For this reason a.c. saturation methods are limited to the detection of defects near to the outside surface.

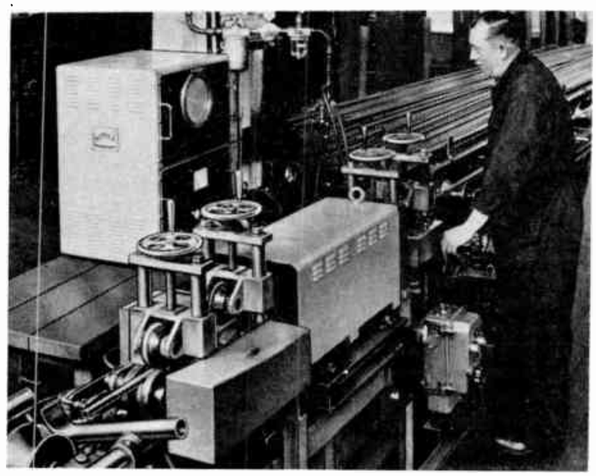
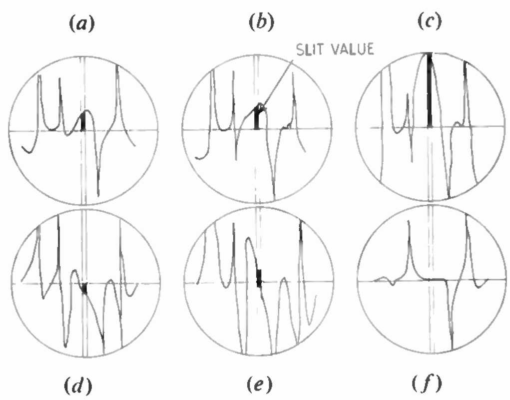


Fig. 4. Eddy-current testing installation using a.c. saturation equipment by Förster.

Equipment using this principle of testing[13] is shown in Fig. 4, together with mechanical handling equipment which propels tubes through a water-cooled test coil at constant speed. Typical curves that appear on the c.r.t. of this equipment are shown in Fig. 5. These are sample traces taken from six tubes during the examination of a batch $\frac{7}{16}$ in o.d. × 17 s.w.g. Traces a, b and c show the effect of serious cracks. Traces d and e are caused by residual stresses introduced when bending the tube by hand. Trace f is typical for sound tubing. Note that there is a distinct difference between crack and stress effect. The whole waveform which represents one complete cycle of test frequency can be moved across the c.r.t. by a control knob until the point of interest lies in the centre of the screen.

A sinusoidal voltage from the control is transformed into a square wave and then differentiated.



(a) cracked tube. (b) cracked tube. (c) cracked tube.
(d) local stress. (e) local stress. (f) sound tube.

Fig. 5. Cathode-ray tube traces for tubes tested on the equipment shown in Fig. 4.

This produces alternate positive and negative pulses; the positive pulses are used to trigger a linear time base and the negative pulses are fed to a 'slit amplifier' which is connected in parallel with the signal voltage to the Y plates of the c.r.t. The 'slit amplifier' is normally muted except for $\frac{1}{20}$th cycle where the amplitude of any signal in the slit is measured and may be monitored automatically. The large peaks on either side of the slit in traces d, e and f are attributed to permeability or dimensional variations and are clearly shown where the excitation field strength passes through points of zero magnitude in the excitation waveform. At these points the effect of initial permeability variations mask the effects from a crack. On the other hand, at points of maximum amplitude in the excitation waveform, saturation reduces the effect of permeability changes but the high current density in the tube intensifies the crack effect.

This type of equipment functions best when the material to be inspected is in a fully annealed condition when internal stresses are reduced to a minimum. The method is confined to the examination of materials exhibiting strong ferromagnetic properties and the same equipment is not suitable for non-magnetic materials or austenitic stainless steels with very low magnetic permeability. The reason for this is that the testing frequency is relatively low to allow separation between dimensions, permeability and crack effects, but the frequency is not high enough to give sufficient electromagnetic coupling between the test coil and non-magnetic specimens. Methods using magneto-mechanical measuring instruments have been suggested as a means of separating stress and crack effects.[14] These are based on the principle that magneto-mechanical measuring instruments or recorders with large inertia do not have a very high frequency response and so the signals due to stresses which have a very high frequency content are not recorded, whereas signals due to cracks have a substantial low frequency content and are recorded.

3.2. *D.C. Saturation*

By superimposing an a.c. field upon a d.c. saturating field the arrangement becomes very satisfactory for inspecting ferromagnetic or stainless steel tubes.[15, 16] With sufficient d.c. field strength the incremental permeability of the material is reduced to unity and the distribution of eddy currents generated by the a.c. field is virtually equivalent to a non-magnetic material having the same dimensions and conductivity.[17] One disadvantage with this method is the necessity for demagnetization after testing. For thin tubes a simple coil fed with a 50 c/s supply is sufficient but for thick tubes or solid rod more elaborate schemes are required;[18] alternatively heat treatment at temperatures above the Curie point will completely demagnetize the material.

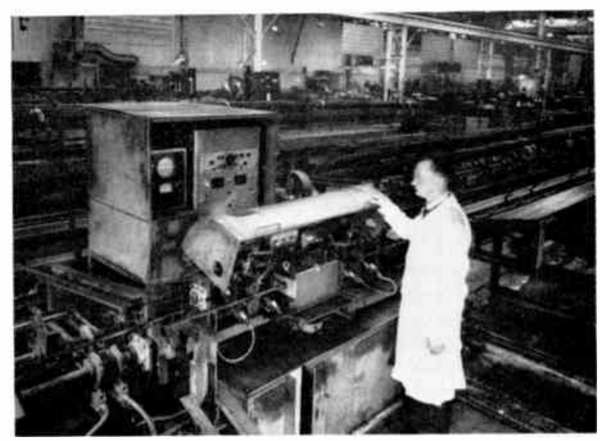


Fig. 6. Fully automatic equipment testing mild steel tubes with d.c. saturation.

Figure 6 shows a fully automatic testing machine designed for inspecting ferromagnetic tubes from $\frac{1}{4}$ in to $1\frac{1}{2}$ in diameter, in lengths up to 60 ft. Tubes are placed on to a rack (r.h. side), selected one at a time by an escapement mechanism and power driven through the test unit (centre) at a constant speed. After passing through the test unit, tubes are automatically sorted into 'passed' and 'reject' storage pallets. The saturating coil is made up of two sections mounted on pole pieces, separated by an air-gap sufficiently wide to allow an eddy current test coil block to be inserted. The d.c. magnetic path is confined to a soft iron yoke completely surrounding the unit which together with soft iron pole pieces reduce the magnetic reluctance. This arrangement allows saturation of the test piece with a much lower power consumption than if an open coil were used.

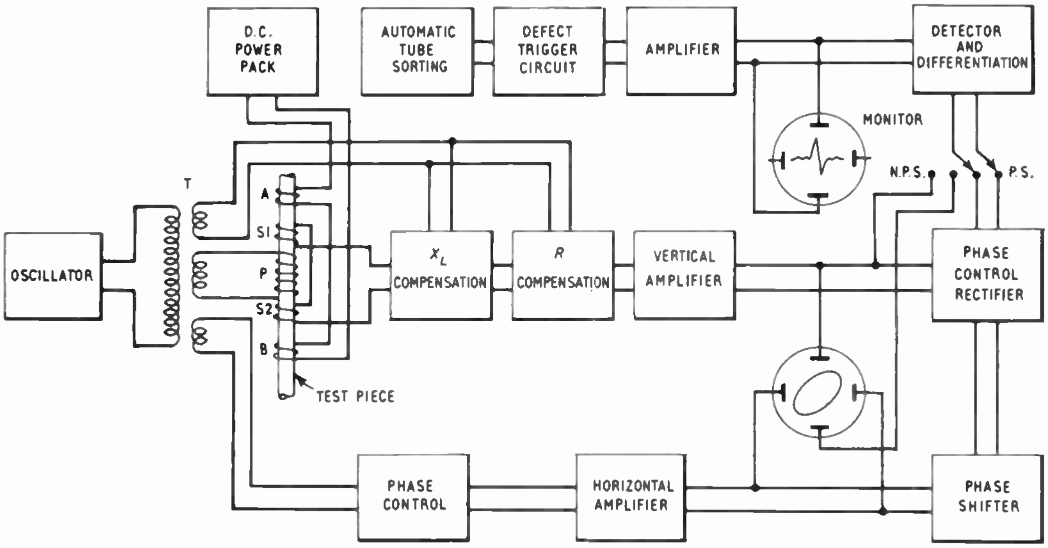


Fig. 7. Block diagram of fully automatic eddy-current testing equipment with d.c. saturation.

'Conductiflux' electronic equipment supplied by Institut Dr. Förster, using the 'ellipse method' can be seen to the left of the photograph with other automatic control equipment. A power driven demagnetizing track is situated behind the electronic equipment.

A block circuit diagram for this equipment is shown in Fig. 7. An oscillator feeds transformer T, having three secondary windings. The centre winding of these three is connected to the exciting winding P of the eddy-current test-coil block. An autocomparison coil system is used, the secondary or detector coils S1 and S2 being connected in anti-phase with each other. With a uniform test piece in position, ideally the resultant output from these detector coils should be zero, but this is never the case because discrepancies occur in the physical dimensions of the coil formers and winding distribution. Any signal due to these causes is neutralized by adding it to another one of equal amplitude and opposite phase by means of XL and R compensation controls, supplied by a secondary winding on transformer T. The third secondary winding on transformer T is used to supply the horizontal plates of the c.r.t. with a reference voltage.

When testing, a defect entering the test coil will result in an output from the secondary windings which is amplified and appears on the vertical plates of the c.r.t., forming a Lissajous pattern of an ellipse with a shape depending upon the signal magnitude and the reference voltage on the horizontal plates. The defect signal is monitored after passing through a phase controlled rectifier, detector or demodulator and differentiation circuit (modulation analysis). The signal is then amplified and applied to the grid of a Schmitt trigger circuit which in turn, fires a thyratron, operating a relay controlling an automatic tube sorting mechanism. Saturation coils A and B are supplied from a stabilized, low voltage d.c. power pack.

4. Signal Processing Techniques

4.1. *Modulation Analysis*

The name 'modulation analysis' has been given to the process of sorting signals out into bandwidths, the frequency range depending upon the nature of discontinuities to be detected and the relative speed between the test coil and workpiece.[19] For the technique to work correctly a constant speed of test is essential. A defect in the part under examination will cause the test frequency to be modulated at a frequency depending upon the rate of change of eddy currents in the part, which in turn is a function of the test speed and coil length. Some discontinuities such as cracks give signals having a short rise time compared with the rise time of signals from some quite harmless metallurgical and dimensional variations. Filter circuits can be designed to suppress signals from some of these tube variations. However, certain defects will pass through such a system without detection because they generate the very same frequency bands which the filters are designed to suppress. These include small or large defects possessing a small depth gradient at their ends.

4.2. *Phase Analysis*

Selecting signals of a given phase as a means of separating defects from dimension or metallurgical differences can only be applied to a very limited degree when testing tubes by a concentric coil. Only one parameter at a time can be suppressed or 'phased out'. In practice, the parameter causing most trouble is suppressed provided sufficient evidence is collected to ensure its identity and to make certain it does not mask a serious defect giving rise to a signal with the same phase. In the majority of cases, variations in i.d. or o.d. are responsible for most signals making up the 'tube noise'.

There are four phase-sensitive methods commonly used, namely the phase-controlled rectifier, the phase discrimination method, the gate method introduced by T. Zuschlag and the unbalance bridge method used in 'Radac' equipment manufactured by Budd Instruments Division, U.S.A. The advantages and disadvantages of these have been dealt with in detail elsewhere.[1, 7, 20, 21, 22]

The block circuit diagram in Fig. 7 shows a switch for phase-sensitive and non-phase-sensitive positions, the selection depending upon the tube o.d. and thickness, e.g. when testing a mild steel tube, $\frac{3}{4}$ in o.d. × 14 s.w.g. thick, at a test frequency of 10 kc/s, signals from internal cracks are 90 deg out of phase with signals from external cracks and therefore phase analysis methods are not justified when used as a criterion for separating cracks from other causes.

In a large tube works several machines may be needed to cover the entire size range and qualities of tube manufactured. The working frequency of such equipment is selected to suit the majority of work in a particular department. In the author's experience equipment with a fixed frequency is more suitable for automatic testing than equipment provided with a switched range of frequencies. The latter is normally sensitive to electrical transients caused by adjacent motors, starters etc., unless elaborate precautions are taken by tuning each stage in the amplifier for each test frequency. This becomes very costly, difficult to maintain and demands greater knowledge and attention from an operator than is desirable. Fixed frequency equipment on the other hand when sharply tuned is not sensitive to transients normally experienced in a tube mill and can be operated by personnel after a short training.

Test frequencies of 1, 10, 100 and 600 kc/s covers most needs for tubes $\frac{1}{64}$ in to $3\frac{1}{2}$ in o.d. in stainless steel or mild steel and intermediate frequencies provide little further advantage.

4.3. *Speed Effect*

Dynamically induced eddy currents, described as 'speed effect' or 'eddy-current drag', are caused by the relative motion between the tube under test and the radial component of flux from the test coil.[23] The effect increases with speed and is found to be more severe in tubes made from good conductivity material such as copper than tubes in mild steel or stainless steel having a much lower conductivity. The effect is constant provided the speed is constant and can easily be balanced out, but it becomes a great handicap when tubes are pushed through the coil with a jerky movement. An application making full use of this effect to advantage has been reported for inspecting shell cases[24] and sheet.[25]

5. Internal Probes

Much attention has been given to devising means of inspecting tubes after they have been built up into units such as condensers, feed water heaters and boilers etc., and checking their condition at intervals during their lifetime.

Portable eddy current testing equipment is now available for doing this job using internal probes of concentric coil construction.[26, 27] With this method of inspection the coil system is more sensitive to bore defects and i.d. variations than o.d. defects and variations. Equipment is normally used with a paper pen recorder to provide a permanent record of the test that may be studied at leisure and comparisons made with previous examinations.

6. Surface Probes

Concentric coils suffer from the disadvantage that very small changes in tube dimensions, conductivity or permeability cause relatively large changes in the electrical properties of the test coil. These variations may be sufficient to mask the effect produced by a small defect even when phase and modulation analysis is used. Surface probes scanning the tube in a helical fashion are not so sensitive to these tube variations which are generally found to occur in a direction along the longitudinal axis, but they are sensitive to circumferential variations which include longitudinal cracks that are continuous or intermittent.[17, 28, 29] A surface defect causes a severe change in the impedance of a pancake coil, but affects the impedance of a concentric coil to a much lesser degree.

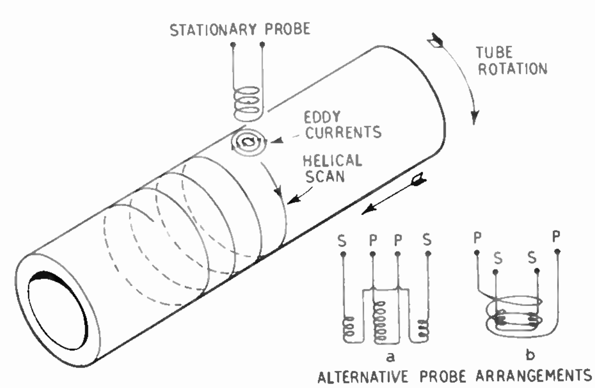


Fig. 8. Inspection by surface probe.

Figure 8 shows the direction of induced eddy currents by a single coil; in this case the probe is kept stationary and the tube rotates beneath it. Two other alternative differential probe arrangements are also shown having separate excitation and detector windings connected back to back which help to reduce the 'lift off' signal normally experienced with single coils. A method for complete suppression of the 'lift off' effect using a single coil in series with a capacitor has been claimed by the Institut Dr. Förster.[1, 30]

For testing long lengths of tube it becomes more convenient to rotate the probe instead of the tube and with speeds up to 2000 rev/min, linear testing speeds in the order of 40 feet per minute are possible.

Modulation analysis can be used with confidence for obvious reasons but if the 'lift off' effect is suppressed by phase analysis some subsurface defects may be missed, depending upon the tube thickness and testing frequency. Although the 'lift off' signal can be suppressed, increase in distance between the probe and tube surface causes a reduction in flaw signal amplitude. To maintain uniform inspection under these conditions an automatic gain control device compensating for probe to tube surface variations becomes necessary.

7. Conclusions

Concentric coil systems can play a very important role in helping to control the quality of tube manufacture provided their shortcomings are clearly understood. Because the speed of eddy-current testing is far higher than any other non-destructive method, it can be introduced into an existing tube making process without upsetting the economy. There is now enough evidence available to state that the concentric-coil method is confined to detecting gross defects in commercial grade tubing but can be used to detect smaller defects in precision tubes where a closer control is kept on dimensions, chemical analysis and metallurgical condition. These factors can vary from one manufacturer to another, or even within the same company, giving rise to a 'noise level' which will limit the degree of resolution that can be reached. Higher resolution to surface defects is possible by scanning the tube with pancake coil surface probes. Difficult mechanical and electrical problems associated with the high peripheral speeds of support bearings and methods of picking up signals from a moving probe are gradually being solved. The present trend is to combine the concentric coil and surface probe techniques to give a more thorough inspection than either one of them can perform alone.

The cost of a non-destructive test may vary from an insignificant amount to a substantial percentage of the final price of a tube. The quantity of tubing to be inspected has a deciding influence upon the method, speed and standard of inspection to be used. When quantities are large, economical advantage can be taken of permanent high-speed testing machines. On the other hand, small quantities of high-quality tubing may need special treatment, individually designed equipment and jigs. The methods employed obviously depend to a great extent upon the ultimate use of the tube or tubular component. For some applications, a visual inspection is good enough, while in other cases, the most advanced techniques are necessary, and may involve supplying each tube with a pen trace recording the extent of tube imperfections.

Ultrasonic tube testing is very slow at the present time and this has retarded its progress, but recent developments with rotating transducers may eventually change the situation. Fundamentally, the ultrasonic method is more sensitive to subsurface defects than the eddy-current method, but until ultrasonic scanning systems are sufficiently developed to cope with large throughputs, the eddy-current method will continue to be the predominating element as an economical non-destructive testing tool on the production line.

At our present level of knowledge inspection of tubes by the eddy current or ultrasonic methods can only be considered on a 'go' or 'no go' basis, by standardizing on artificial defects having various shapes in the form of drilled holes or milled slots.

8. Acknowledgments

The author wishes to thank his colleagues for their valuable help and the Directors of Accles & Pollock Ltd., for permission to publish this paper.

9. References

1. R. C. Macmaster, "Non-destructive Testing Handbook", Vol. 2, (Ronald Press, New York, 1959).
2. F. Förster, "Developments in non-destructive testing of materials with electromagnetic methods", *Solus-Schall News*, No. 25, October 1962.
3. H. L. Garbarino and H. N. Nerwin, "Progress in technology and equipment for eddy current inspection", *Non-destr. Test. (U.S.)*, p. 229, July/August 1959.
4. P. Graneau and S. A. Swan, "Electromagnetic fault detection on non-ferrous pipes", *J. Electronics & Control*, **8**, No. 2, p. 127, 1960.
5. R. Hochschild, "Mathematical Foundations of Non-destructive Testing by Eddy Current Methods", U.S.A.E.C. Report No. NYO-3576, 1953.
6. R. Hochschild, "New German Instruments for Non-destructive Testing and their Applications to Commission Problems", U.S.A.E.C. Report No. NYO-3577, 1952.
7. R. Hochschild, "Electromagnetic Methods of Testing Metals", "Progress in Non-Destructive Testing", Vol. 1, pp. 57–109. Ed. E. G. Stanford and J. H. Fearon (Heywood, London, 1958).

8. C. W. Clapp, "Detecting tramp metal in logs and iron ore", *Electronics*, **24**, p. 88, March 1951.
9. F. R. Bareham, "Choice of frequency for eddy current tube testing", *Brit. J. Appl. Phys.*, **11**, p. 218, June 1960.
10. B. H. Robinson, "Frequency selection for eddy current testing", *Non-destr. Test.* (*U.S.*), p. 36, January/February 1958.
11. F. Förster and H. Breitfeld "Quantitative flaw detection in metals by the slip-over coil", *Z. Metallkunde*, **45**, No. 4, p. 188, 1954, (H. Brutcher Translation No. 3329).
12. F. Förster, K. Sprungmann, A. Michalski, W. Kock, "Electromagnetic methods of non-destructive testing", *Techn. Mitt.* **50**, No. 4, p. 162, 1957, (H. Brutcher Translation No. 3956).
13. F. Förster, "Magnetic-inductive crack detection in steel", *Z. Metallkunde*, **45**, No. 4, p. 221, 1954, (H. Brutcher Translation No. 3401).
14. K. Matthaes, "Influence of residual stresses on detection of flaws in steel tubing by electromagnetic means", *Z. Metallkunde*, **44**, p. 473, 1953 (H. Brutcher Translation No. 4332).
15. "New developments in non-destructive testing of precision steel tubes", *Instrument Practice*, **15**, p. 553, May 1961.
16. H. C. Knerr and C. Farrow, "Method and Apparatus for Testing Metal Articles for Defects", U.S.A. Pat. No. 2,065,379, February 1939.
17. W. H. Baker, "Detection of flaws in steel tube or bar with a rotating coil in a magnetic saturating field", *J. Brit.I.R.E.*, **24**, No. 1, p. 77, July 1962.
18. C. Farrow, "Method and Apparatus for Demagnetizing Tubing", U.S.A. Pat. No. 2,460,684, May 1943.
19. R. Hochschild, "Modulation analysis—a new instrument technique in eddy current testing". *Non-destr. Test.* (U.S.), p. 323, September/October 1960.
20. D. L. Davies, "A fast-acting phase conscious indicator" *Electronic Engineering*, **28**, p. 383, September 1956.
21. R. Chidambaram and S. Krishnan, "Diode phase-sensitive detectors with load", *Electronic Engineering*, **31**, p. 613, October 1959.
22. T. Zuschlag, "Magnetic Analysis Inspection in the Steel Industry", A.S.T.M. Publication No. 85, 1949.
23. P. Graneau, "The effect of sample movement in fault detection using eddy currents", *Proc. Instn Elect. Engrs*, **106B**, p. 417, 1959 (I.E.E. Paper No. 2945M, July 1959).
24. "An Eddy Current Method for Inspection of Shell Cases". U.S. Naval Report No. 0-2282, April 1944.
25. R. Gunn, "Apparatus and Method for Detecting Defects in Metallic Objects", U.S.A. Pat. No. 2,255,053, September 1941.
26. J. B. C. Robinson and L. T. Perriam, "The inspection of tubes for flaws and variations in thickness", *J. Inst. Metals*, **86**, p. 316, 1957–58.
27. P. L. Datt and T. R. Schmidt, "Some applications of eddy current testing in the petroleum industry", *Non-destr. Test.* (*U.S.*), p. 335, September/October 1962.
28. J. W. Allen and R. B. Oliver, "Inspection of small-diameter tubing by eddy current methods", *Non-destr. Test.* (*U.S.*), p. 104, March/April 1957.
29. H. F. Ammarell and R. B. Moyer, "Surface defect detection", *Wire and Wire Products*, **36**, No. 11, November 1961.
30. "Eddy current flaw detector is sensitive and versatile", *Metalworking Production*, p. 73, 7th November 1962.

Manuscript received by the Institution on 1st February 1963. (*Paper No. 860.*)

U.D.C. 621.387.42

A New Poly-anode Counting Tube, the 'Polyatron'†

By

Professor Y. HATTA,‡

H. MASE‡

AND

M. SUGAWARA‡

Summary: A new type of gas-filled counting tube has been developed. This tube contains a common cathode with many anodes around it, differing from the dekatron of conventional type. It may be driven by small input pulses of amplitude about one-half as large as that of a conventional type. It can drive the number display tube without any intermediate device, and is expected to have higher reliability. The paper describes the static and dynamic characteristics and the application of this new counting tube.

1. Introduction

The gas-filled counting tube, known as the dekatron or stepping tube, is one of the simplest decade counters[1,2]. This type of tube has a common anode with many cathodes surrounding it. A glow on one of the cathodes is transferred to the next by the application of a negative pulse to guides, which are located between adjacent cathodes. A number of negative input pulses thus corresponds to that number of transferences of the glow.

There are many kinds of these tubes, but all of them have similar electrode structure, and are of the so-called poly-cathode type.

If it were possible, however, to realize a new type of dekatron having a common cathode and many anodes around it, in other words, a dekatron of inverse polarity, the following advantages over the dekatron of normal type could be expected.

(*a*) The possibility of obtaining higher reliability of operation, since the effect of sputtering of the cathode material which destroys the uniformity of the emission characteristics among many cathodes and reduces the reliability, is removed in this new electrode structure.

(*b*) The possibility of driving a number display tube by using the new dekatron without any intermediate device, because number display tubes are of the poly-cathode type and can be connected in series with this new dekatron.

(*c*) The new dekatron would be driven by a smaller pulse amplitude, because the anode fall is generally smaller than the cathode fall.

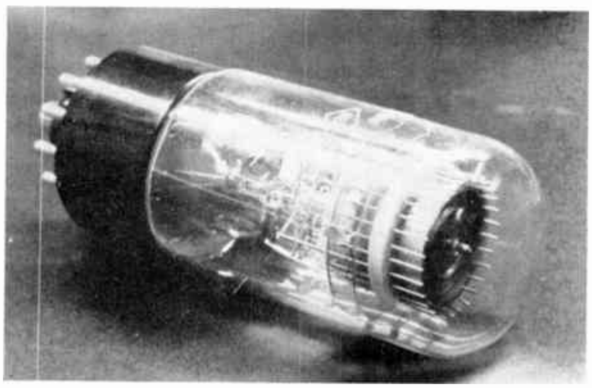
In order to corroborate these ideas, the authors have developed a new dekatron, the poly-anode dekatron, which has been given the name of 'Polyatron'.† This paper describes the characteristics and the application of the poly-anode dekatron.

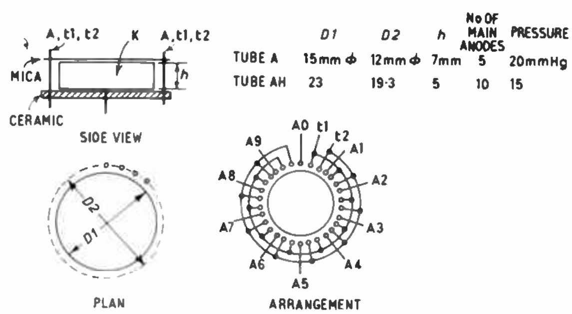
† Japanese Pat. Pend.

‡ Department of Electronic Engineering, Tohoku University, Sendai, Japan.

2. Description of the Tube

The tube consists of a cylindrical cathode surrounded by thirty rod-like anodes. The electrode configurations are shown in Fig. 1.





	D1	D2	h	No OF MAIN ANODES	PRESSURE
TUBE A	15mm ɸ	12mm ɸ	7mm	5	20mmHg
TUBE AH	23	19·3	5	10	15

Fig. 1. Construction and arrangement of electrodes of the poly-anode dekatron. (A0–A9 main anode; t1, t2 transfer anode; K cylindrical cathode.)

Ten main anodes (A0 to A9) are individually brought out to ten base pins. Twenty transfer anodes (or guides) are connected in two groups (t1 and t2) as in the case of the guides of a conventional double pulse dekatron, and the tube is driven by a positive consecutive double pulse impressed on the transfer anodes.

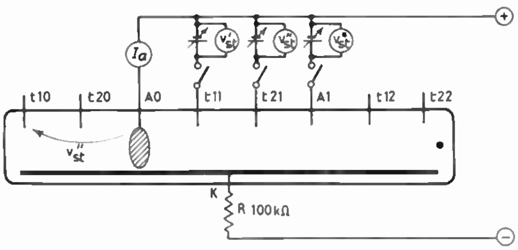


Fig. 2. Measurement of static stepping voltages, v'_{st}, v''_{st} and v^*_{st}.

Dimensions of the electrodes are determined so that the stable dekatron operation may be obtained when the anode current is 1·0–2·0 mA. The tube is filled with the mixture of 90% argon and 10% hydrogen. In order to examine the effect of hydrogen mixing, the 'Tube A' is filled with 100% argon. The test tube filled with argon and hydrogen is named 'Tube AH'.

3. Static Characteristics of the Tube

The static characteristics of the tube can be measured by using two adjacent main anodes and two transfer anodes between them connected as shown in Fig. 2.

The anode current I_a flows from A0 to K. In this case, when the applied potential of t11 with respect to A0 is increased, I_a will step from A0–K to t11–K space at a certain critical potential difference between A0 and t11, which is called[3] the static forward stepping voltage v'_{st}. Similarly, the stepping voltage to t21 can be measured and is called the static backward stepping voltage v''_{st} and the stepping voltage to A1 is denoted as v^*_{st}. The three stepping voltages, v'_{st}, v''_{st} and v^*_{st} vary with I_a and the measured characteristics for 'Tube A' are shown in Fig. 3(*a*).

In Fig. 3(*a*), v'_{st} is smaller than v''_{st} and v^*_{st} in the region of $I_a < 2{\cdot}5$m A, when it can be expected to realize the normal operation as a stepping tube or dekatron. As shown on the v'_{st} curve in Fig. 3(*a*) two kinds of stepping characteristic can be found, i.e. the region dependent on I_a (Region I) and an independent region (Region II). According to those definitions, v''_{st} and v^*_{st} of Fig. 3(*a*) belong entirely to Region II.

In Region I, I_a is relatively large and t11 is immersed in plasma (this is verified by probe measurement). The electron current flowing into t11 increases with increased applied voltage as in the case of probe measurement. Further increase of the applied voltage can easily break down the space between t11 and plasma, and the main anode current is then shared by A0 and t11. By definition the stepping of I_a from A0 to t11 is finished when the current to A0 becomes zero. Therefore, v'_{st} is equal to the floating potential of t11 with respect to A0. Since the floating potential is nearly independent of the main current, variation of v'_{st} with I_a in Region I is small.

In Region II, the anode, to which I_a steps, is not immersed in plasma but in a space charge of relatively low density which is verified by checking the 'Debye length'. Therefore, the stepping of I_a to another anode can preferably be understood as the breakdown between this anode and the common cathode. In general, if the electric field is distorted by the presence of a space charge, the breakdown voltage decreases when the gas pressure is not so small. As the discharge of A0–K space distorts the electric field of its surrounding space, the breakdown voltage between another anode and the common cathode (V_{st}) is reduced by this field distortion, as shown in Fig. 3(*b*). Thus, in Region II, v_{st} $(=V_{st}-V_m)$ decreases with I_a.

Moreover, v''_{st} is affected by the potential or the bias voltage of t11 (E_t), as shown in Fig. 4, and can be explained as follows.

If the potential of t11 is lower than the plasma potential, a positive ion sheath is formed around t11.

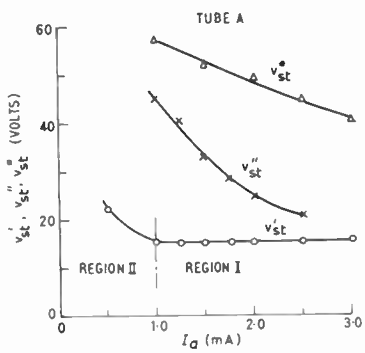
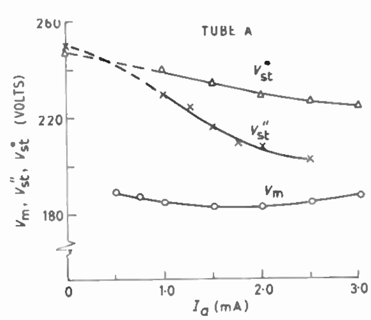


Fig. 3. (*a*) The static stepping voltages, v'_{st}, v''_{st} and v^*_{st} vs. anode current I_a for Tube A.



(*b*) Breakdown potential of t21 and A1, V''_{st} and V^*_{st} vs. anode current I_a for Tube A. V_m is the maintaining potential and so $V^*_{st} - V_m$ gives v^*_{st}.

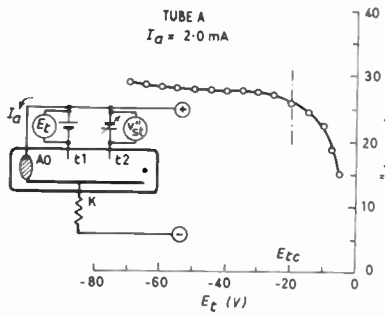


Fig. 4. Static backward stepping voltage v_{st}'' vs. negative bias E_t of t11 for Tube A. To the right of E_{tc}, v_{st}'' is nearly constant while it varies steeply to the left of E_{tc}.

This sheath spreads up to the neighbouring space of cathode and controls v_{st}'' just as in the case of grid control of a thyratron. When E_t is equal to E_{tc} shown in Fig. 4, the calculated thickness of the positive ion sheath is nearly equal to the distance between t11 and K. Therefore when $|E_t| > |E_{tc}|$, t21–K space is not influenced by the discharge of A0–K space owing to the presence of the ion sheath around t11, resulting in nearly constant v_{st}''. On the other hand, when $|E_t| < |E_{tc}|$, the thickness is smaller than the anode-cathode distance and so v_{st}'' is easily affected by I_a.

The step voltage v_{st}^* corresponds to the maximum output voltage to be obtained.

In order to achieve stable operation as a counting tube, it is desirable that v_{st}' is as small as possible and v_{st}'' or v_{st}^* is as large as possible. Accordingly, v_{st}' must be chosen to be in Region I at a required current. To obtain large v_{st}'' and v_{st}^*, a contracted density distribution of electrons and ions and a large difference between the breakdown and the maintaining potential are desirable.

In the case of Tube AH, v_{st}^* is larger than that of Tube A, as shown in Fig. 5. Thus the static characteristics are satisfactory.

4. Dynamic Characteristics of the Tube

To determine the dynamic characteristic, the minimum and the maximum amplitudes of the driving pulse which are necessary to obtain the correct counting operation are plotted against the frequency f of the input pulse. This is obtained by using the circuit and input pulse waveforms shown in Fig. 6.

The result with Tube A is given in Fig. 7, in which the minimum and the maximum amplitudes of the driving pulse (v_d' and v_d'') are expressed by the effective pulse height ($v_d = v_p - |E_t|$), where E_t is the negative bias voltage of transfer anodes with respect to the index anode and v_p is the true amplitude of the input pulse. Between those two limits, stable operation can be attained.

When pulse width (τ_1 of Fig. 6) is kept constant, v_d' is almost independent of f, but varies with the pulse width, as shown in Fig. 8. From those characteristics, it can be understood that the physical meaning of τ_1 which is necessary for the stable operation is the time lag of breakdown. When f is increased, there appears to be a sudden decrease of v_d'' as shown in Fig. 7, which can be attributed to the effect of residual ions

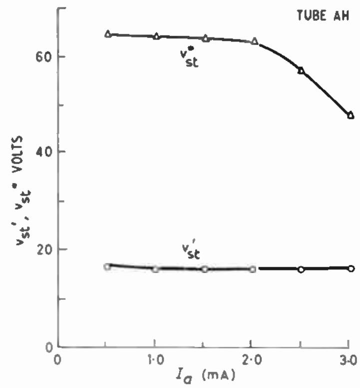


Fig. 5. Static stepping voltages of Tube AH, v_{st}' and v_{st}^* vs. anode current I_a.

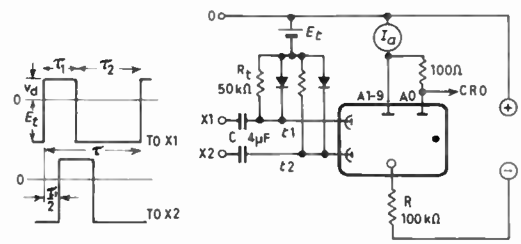


Fig. 6. Circuit for measuring the dynamic characteristic and the input waveforms.

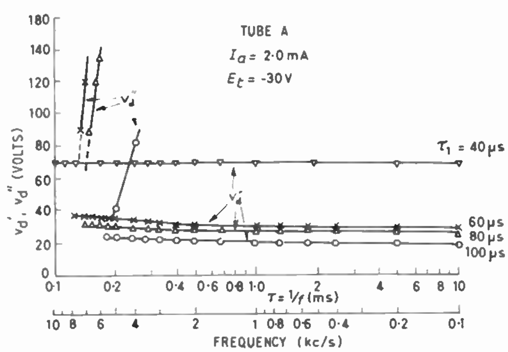


Fig. 7. Dynamic characteristic of tube for various pulse widths when anode current I_a is 2·0 mA and negative bias voltage E_t is fixed at $-$ 30 V. v_d' and v_d'' are the minimum and the maximum pulse heights for correct operation respectively, which are expressed by the effective pulse height ($v_d = v_p - |E_t|$).

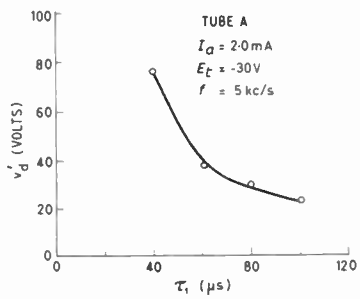


Fig. 8. The minimum pulse amplitude for correct operation, v'_d vs. pulse width τ_1.

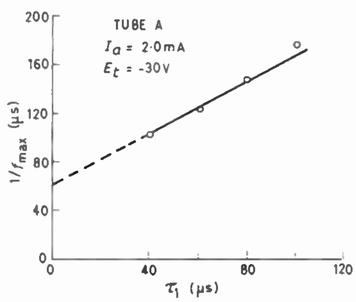


Fig. 9. The maximum counting frequency f_{max} vs. pulse width τ_1.

of the foregoing discharge. If the residual ions of the foregoing discharge cannot sufficiently decay during τ_2^* of Fig. 6, the glow may step backward within $\tau_1/2$. Since such effect of residual ions increases when f is increased, v''_d decreases with increased f.

The region for the correct operation is limited by this drooping characteristic of v''_d, and the maximum counting frequency (f_{max}) corresponds to a cross-over point where $v'_d = v''_d$. If f_{max} is limited by a de-ionization time (τ_2^*) during which the concentration of residual ions decreases to a certain value, f_{max} can be expressed as

$$\frac{1}{f_{max}} = \tau_1 + \tau_2^*$$

Figure 9 shows that the relation between τ_1 and $1/f_{max}$ is linear which means that τ_2^* is constant (60 μs) during the experiment.

The de-ionization time can be determined by measuring the frequency of the relaxation oscillation of the tube and was found to be 65 μs. This numerical value agrees with that of τ_2^* obtained from Fig. 9, and supports the expression for f_{max}. When $|E_t|$ is large, f_{max} tends to rise, because the negative potential of the transfer anodes accelerates the de-ionization.

The dynamic characteristics of Tube AH are shown in Fig. 10, from which it may be seen that f_{max} for this tube is about 25 kc/s. Comparing this with Tube A, it is clear that the effect of hydrogen mixing shortens both the time lag of breakdown and the de-ionization time.

5. Application of the Tube

This tube can be used not only as a counting device, but also as a driving device for direct operation of a number display tube. The basic circuit is illustrated in Fig. 11.

As the number display tube T1 is the poly-cathode type, it can be connected in series with this poly-anode tube T2. It is impossible to apply a larger potential difference than v_{st}^* between a glowing anode and an adjacent anode of T2. On the other hand, the number display tube requires a pre-bias voltage of some +40 V for satisfactory operation.[4, 5] Therefore the voltage between a glowing anode and an adjacent anode of T2 ($V_{aa'}$) or the voltage between a glowing cathode and an adjacent cathode of T1 ($V_{kk'}$), must be fixed within a recommended range, i.e.,

$$+40\ \text{V} < V_{aa'} (= V_{kk'}) < V_{st}^*$$

In some cases, cathodes of T1 (K0–K9) are connected to each other through the resistor R′, as shown in

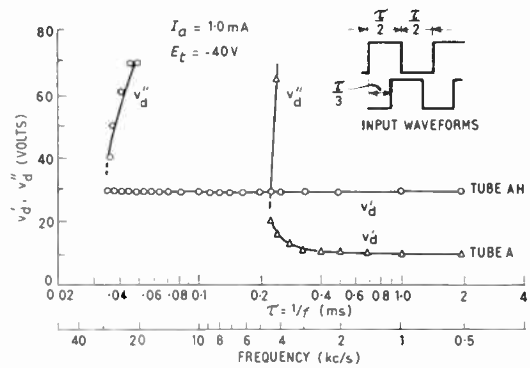


Fig. 10. Dynamic characteristics of Tube AH and Tube A, when I_a is 1·0 mA and E_t is fixed at −40 V. The input pulses are shown. v'_d and v''_d are the minimum and the maximum pulse heights for correct operation respectively, which are expressed by the effective pulse height ($v_d = v_p - E_t$).

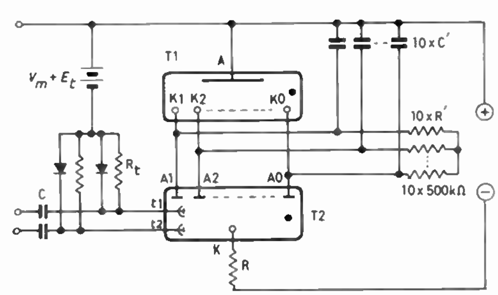


Fig. 11. Basic circuit for direct operation of the number display tube.

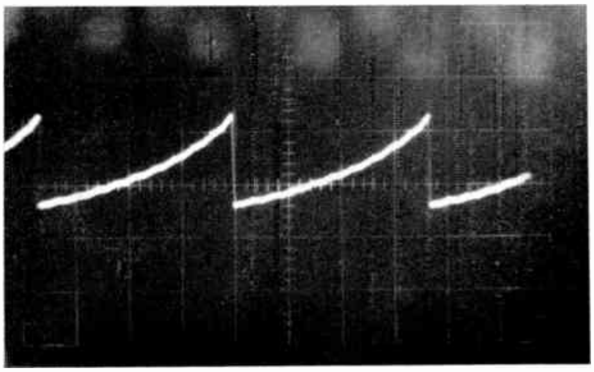


Fig. 12. Anode potential waveform of Tube AH.

Fig. 11. Then, $V_{aa'} = V_{kk'}$ will be fixed to $2R'I_a'$, where I_a' is the ion current flowing into an adjacent cathode.

Capacitors C′ are added in order to avoid rapid variation of the load of T2.[4] The capacitance is determined to keep the time-constant $C'R'$ at about $10/f$ seconds at operating frequency f. The waveform of the anode potential (or output voltage) is shown in Fig. 12.

Since this circuit does not require any intermediate device, the circuit connection becomes very simple compared with those of the other driving methods.[4]

Recently D. Reaney[5] has developed a new type of dekatron for the same purpose, but it is believed that the authors' method has two merits over that of Reaney, namely:

(a) Electrode configuration is simple,

(b) The current of the dekatron is equal to that of the number display tube, while in Reaney's method, the former is 1·5 times as large as the latter.

6. Conclusion

The poly-anode dekatron (Polyatron) is a tube developed for realizing the three objects mentioned in the introduction. The tube can be employed not only as a counting tube, but also as a driving device for a number display tube. The amplitude of the input pulse for the correct operation is about one-half as large as that of conventional dekatron. In order to drive the number display tube, the tube does not require the additional control elements that are needed with the conventional dekatron, and the circuit is very simple.

The Polyatron is expected to have higher reliability than the conventional dekatron. The running tests to prove this are being continued.

The explanation of static and dynamic characteristics gives useful data for design of this tube.

7. Acknowledgments

This work is supported by the Fund for Scientific Research of Omi Kenshi Spinning Co. and the authors express their thanks for this financial aid. They wish to thank Dr. H. Nishio and Dr. H. Kobayashi of the Nippon Electric Co. who helped this work by making tubes for the experiments.

8. References

1. R. C. Bacon and J. R. Pollard, "The dekatron—a new cold cathode counting tube", *Electronic Engng*, **22**, p. 173, 1950.
2. J. H. L. McAsulan and K. J. Brimley, "Polycathode counter tube applications", *Electronics*, **26**, p. 138, November 1953.
3. Y. Hatta, N. Hanzawa and M. Sugawara, "Static characteristics of the single pulse dekatron", *ETJ Japan*, **2**, p. 82, 1956.
4. G. Higgins, "A gas-filled glow discharge character display tube", *J.Brit.I.R.E.*, **22**, p. 133, 1961.
5. D. Reaney, "A new dekatron for direct operation of digitrons", *Electronic Engng*, **34**, p. 372, 1962.

Manuscript first received by the Institution on 26th March 1963 and in final form on 24th June 1963. (Paper No. 861.)

INSTITUTION NOTICES

Higher National Certificates and Diplomas in Electrical and Electronic Engineering

The Ministry of Education announced on 31st October 1963, in a circular to all technical colleges in England and Wales, that the Ministry has considered the position of the existing Higher National Certificates and Diplomas in Electrical Engineering and that:

"After consultations with the Institution of Electrical Engineers and the British Institution of Radio Engineers, it has been agreed that the title of these certificates and diplomas should be widened to include Electronic Engineering and that a new Joint Committee should be established consisting of:

three representatives of the Ministry of Education; three representatives of the Institution of Electrical Engineers; three representatives of the British Institution of Radio Engineers; and three members appointed on the joint nomination of the Association of Technical Institutions, the Association of Teachers and Technical Institutions, and the Association of Principals of Technical Institutions."

The new Committee is being set up at once and its function will be to administer the rules of the old scheme and draft and administer the rules for the new scheme. It is envisaged that the last course in the old scheme should start in September 1964 although some colleges may be in a position to offer the new Certificate or Diploma course in that year.

Annual Report of the Council

The Annual Report of the Council of the Institution for the year ended 31st March 1963 is published in the November issue of the *Proceedings of the Brit.I.R.E.* All members in Great Britain have been sent copies; members overseas may receive a copy of this issue of the *Proceedings* free of charge on application to the Secretary of the Institution, 9 Bedford Square, London, W.C.1.

Change of Name of the I.R.E. of Australia

At an extraordinary general meeting of The Institution of Radio Engineers Australia held in Sydney on 30th October 1963, it was decided unanimously to change the name of the Institution to "The Institution of Radio and Electronics Engineers Australia". The new title will be effective from 1st January 1964.

U.K. Delegation to Geneva Radio Conference

A delegation consisting of representatives of various Government Departments and led by Captain C. F. Booth, C.B.E., a former Deputy Engineer-in-Chief of the General Post Office, represented the United Kingdom at the Space Radio-communication Conference which opened in Geneva on 7th October.

The Conference, organized by the International Telecommunication Union, had, as its main task, the allocation of frequencies for space research and radio astronomy.

The frequency bands best suited for space communication purposes lie in a part of the spectrum (1000 to 10 000 Mc/s) which is already heavily committed for existing or planned conventional radio services. One of the main problems before the Conference was to agree on conditions that will ensure that land and space radio services will not interfere with one another. The radio regulations adopted by the Geneva Radio Conference in 1959, are to be revised to take account of all the decisions of the present Conference, including the necessary technical measures to be adopted by Administrations.

Frequency allocation is a necessary first step before the establishment of space communication services can be considered and the International Telecommunication Union restricted the Conference agenda to frequency matters.

F.E.A.N.I.

The Fédération Européenne d'Association Nationales d'Ingénieurs, which held its Fourth Congress at Munich on 17th June this year, has initiated the preparation of a European register of professional engineers.

The F.E.A.N.I. was formed in 1951 to try to interrelate the education and training of engineers in different countries in Europe and, in particular, to deal with matters concerned with university and non-university trained engineers. By its system of national representation F.E.A.N.I. can now be said to represent the interests of 350 000 professional engineers.

At the present moment the United Kingdom is not represented, but observers were present at the Congress. Since representation is through national bodies, if Great Britain were to join, it would be through a body such as the Engineering Institutions Joint Council rather than through individual Institutions.

Conference on "The Peaceful Uses of Atomic Energy"

H.M. Government has accepted the invitation of the United Nations to participate in the Third International Conference on "The Peaceful Uses of Atomic Energy" to be held in Geneva from 31st August to 9th September 1964. The main theme of the Conference will be new developments in power reactor technology but sessions are also proposed on the new developments in controlled thermo-nuclear reactions, applications of radio-isotopes, research reactors and isotope separation. Full details of the Conference may be obtained from: Geneva Conference Secretariat, 11 Charles II Street, London, S.W.1.

U.D.C. 621.37/8: 665.58

Electronic Instrumentation in Petroleum Refineries

By

W. H. TOPHAM, B.Sc.†

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

Summary: Electronic process control systems have overcome the transmission limitations of conventional pneumatic equipment for centralized control, and have facilitated the application of electronic data logging and computational procedures. Product quality is the key parameter in refinery operations and continuous analysers have been developed to overcome the limitations of conventional laboratory testing. The design and performance criteria for process monitors are outlined, and instruments are described for the on-line measurement of specific gravity, boiling range, colour and composition. Continuous analysers can be used for process control and line blending, and similar analytical techniques could be applied to measure the quality of products in storage vessels. Electronic instruments are used for tank level and temperature gauging, flow metering and blending. New instrumental techniques are supplementing conventional methods for plant inspection and maintenance.

1. Introduction

The petroleum industry has always appreciated the capabilities and potential benefits of instrumentation, and oil refining is among the most highly instrumented and automated of modern industries. Within the last two decades, electronics has become the dominant technology in instrumentation and control, and this is reflected in its applications in refineries.

A refinery converts crude oil into a wide range of fuels and lubricants, by means of such processes as distillation, cracking and reforming, extraction, sweetening and blending. A large modern refinery may process 10 million tons of crude oil every year, equivalent to a continuous throughput of 5000 gallons (roughly two road tankers) every minute. To handle this enormous quantity, plants have to operate day and night throughout the year, an important point to bear in mind when considering the reliability demanded of on-line instruments.

Instrumentation can be considered under the following groups of applications.

(*a*) Measurement and control of process operating conditions or environment (e.g. temperature, pressure, level and flow).

(*b*) Measurement and control of product quality.

(*c*) Quantity measurements and blending.

(*d*) Plant inspection and maintenance.

From the point of view of increasing productivity, the author considers that instrumentation for the measurement and control of product quality is potentially the most important of the applications listed above, and therefore the paper deals primarily with this subject. Other applications are described only briefly, to illustrate the diversity of electronic instrumentation in refineries.

2. Process Control Equipment

Until as recently as five years ago, oil refineries had little alternative but to use pneumatic systems to control process operating conditions, and even today, pneumatic equipment is far from obsolescent. The most common primary measuring elements are thermocouples for temperature, orifice meters for flow, floats and displacers for level, and bourdon tubes for pressure. Each measured variable is converted into an equivalent air pressure between 3 and 15 lb/in^2 and transmitted to a two- or three-term controller located in the plant control room. The pneumatic output signal controls the position of a diaphragm-operated valve to regulate the flow of oil, fuel, steam or water. Air-operated systems have the advantages of safety, simplicity, standard signal level, high power output and reliability, but are restricted to transmission distances below about 500 feet.

In recent years, the trend in refinery design has been to centralize the control of a group of plants, to reduce initial capital cost and provide more efficient operation with fewer personnel. Centralization may involve long signal transmission lines for which pneumatic systems are unsuitable, and the availability of all signals at one central location facilitates the application of data processing techniques. These requirements have stimulated the development of electronic process control systems.

† BP Refinery (Kent) Ltd., Isle of Grain, Rochester, Kent.

Electronic systems generally make use of the same types of primary measuring elements as pneumatic systems, with the mechanical movement converted by means of a differential transformer, pot coil, etc, into an equivalent low-level electrical signal, e.g. 10 to 50 mA d.c. or $-\frac{1}{2}$ to $+\frac{1}{2}$ V a.c. Strain gauges have also been applied recently for primary process measurements, and would appear to have some attractive features.

The controller circuits for proportional, integral and derivative action may use conventional electronic tubes, but transistors and other solid state elements are also being widely used.

The applications and achievements of high speed electronic data logging and computer control have been described in several excellent papers presented at the Brit.I.R.E. 1963 Convention and elsewhere.[1, 2] However, it should not be assumed that computational devices must necessarily be electronic. The conventional pneumatic process controller is itself a simple computer, and many other pneumatic devices are available for simple calculations such as summation, square root extraction, etc.

3. Measurement and Control of Product Quality

The quality of each refinery product must comply with an agreed specification, and a great deal of testing is necessary to satisfy this requirement. To achieve maximum yields and the most economic operation, each product should be precisely to the required specification. In other words, the quality 'give-away' should be reduced to the minimum. How far this can be realized in practice depends on the accuracy and frequency of quality measurements.

The traditional, and still the most widely used, method of controlling the quality of process streams is by inference from operating conditions (e.g. temperature, pressure and flow) based on past experience and supplemented by routine laboratory analyses. This method can produce satisfactory specification products but not necessarily at the best yield or with minimum quality 'give-away'.

To improve the precision and speed of laboratory testing, a variety of semi-automatic analytical instruments have been developed during the last decade, using electronic techniques to simulate many of the routine, empirical test procedures.[3, 4] However, instrumentation of this type, although successfully applied in its own sphere, can do little to overcome the fundamental deficiency of laboratory testing for process control, which is the time lag between sampling and providing the result on which to base corrective action. Also, the one analysis represents conditions only at the time the sample was taken, and since product quality from a continuous process may be fluctuating, control based on periodic sampling is not conducive to efficient operation.

3.1. *Continuous Quality Analysers*

To overcome the limitations of inferential control and laboratory testing, analytical instruments have been developed for continuous on-line operation on plants. These instruments, often referred to as process monitors, provide the plant operator with an up-to-date record of product quality. Control action can be taken as a change occurs, to maintain quality continuously at the desired level, and significant improvements in plant efficiency and economy may be obtained. In some cases, the analyser may be incorporated in a closed-loop to control the product quality automatically.[5] Product quality is generally the key parameter in the operation of refinery processes, therefore the development of continuous analysers is an important stage in the successful and effective application of computer control.

The electronic techniques used in quality analysers are in general the same as those used for automatic laboratory instruments, but there the resemblance ends. Process monitors, in common with other on-line instruments, must satisfy the following conditions:

(*a*) Electrically safe for operation in potentially explosive atmospheres.

(*b*) Fail-safe under all fault conditions.

(*c*) Completely automatic, for continuous, unattended operation for several months.

(*d*) Remote transmission and recording of the measured value.

(*e*) Ruggedly constructed for reliable operation in exposed locations.

(*f*) Simple to install, and easily maintained and calibrated.

(*g*) Consistent accuracy and repeatability over long periods.

Electrical safety can be achieved by using equipment which is either intrinsically safe (B.S. 1259), flameproof (B.S. 229), explosion proof (American electrical code) or air-purged. The last named has the added advantages of simplicity, versatility and providing a clean, dry internal atmosphere conducive to reliable operation.

Reliability depends to a large extent on good design, but can certainly be improved by using the simplest possible instrument to perform the required analysis. Unfortunately, this approach may not always be compatible with instrument manufacturing practice and marketing requirements, and a compromise may have to be accepted between reliability and versatility. Long-term precision can probably best be achieved by employing null-balance techniques wherever possible, to reduce the effects of changes in the characteristics of electronic components.

The criteria for process analysers are more stringent than those for laboratory instruments, and can only be satisfied by design and engineering of a high order. Nevertheless, many such instruments are available and some of these are described below.

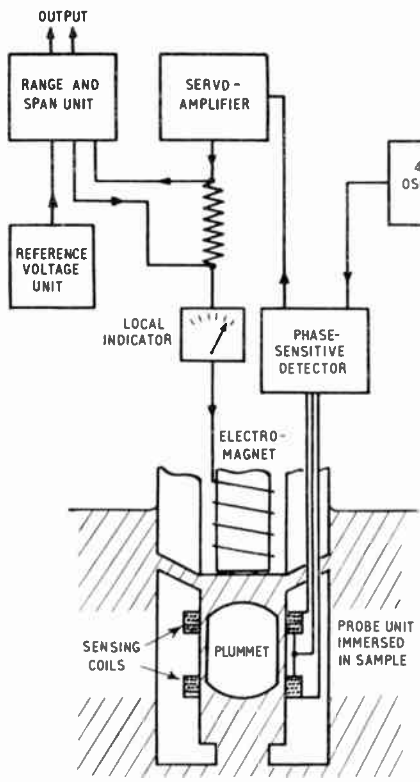


Fig. 1. Diagram of specific gravity monitor with servo-suspended plummet. (General Communication Company.)

3.1.1. Specific gravity monitor

Specific gravity (or density) is used in refineries either to convert volumetric flow measurements into mass units, or to indicate some other property which cannot easily be measured. The standard laboratory method, using a hydrometer and thermometer, is probably one of the simplest and quickest, yet most accurate (0·05%) of all petroleum tests. Figure 1 shows the principle of operation of a specific gravity monitor used successfully at Kent Refinery on a wide range of products of different viscosities. A soft-iron plummet immersed in the flowing sample is servo-suspended beneath an electromagnet. Any change in the position of the plummet, caused by a change in sample gravity, is detected by two concentric sensing coils fed with anti-phase signals at 490 kc/s. The out-of-balance output from the phase-sensitive detector drives an amplifier which changes the current through the electromagnet and restores the plummet to its original position. The current change is proportional to the change in specific gravity, and the instrument can detect s.g. changes of 0·02%, with good long-term stability and precision.

3.1.2. Distillation point monitor

Distillation is the most important and widely used process in oil refining for separation of oil mixtures into products having a specified boiling range (e.g. motor spirit, kerosene and Diesel fuel). The standard laboratory test for measuring distillation range consists of boiling 100 ml of the sample in a glass flask over a gas burner and noting on a mercury thermometer the vapour temperatures corresponding to the first drop distilled, then each 10% of distillate, and when the last drop evaporates. Complex electronic laboratory instruments have been designed[3] which can reproduce automatically the rigorously defined standard test procedure, but such instruments are both unnecessary and unsuited for continuous operation on plants.

Recently, a plant monitor to record selected distillation points (e.g. 10% and 90% points) on up to six process streams in sequence has been developed by the author's organization.[6] The instrument, illustrated in Fig. 2 is similar in operation to the standard test, but is completely automatic and ruggedly built for

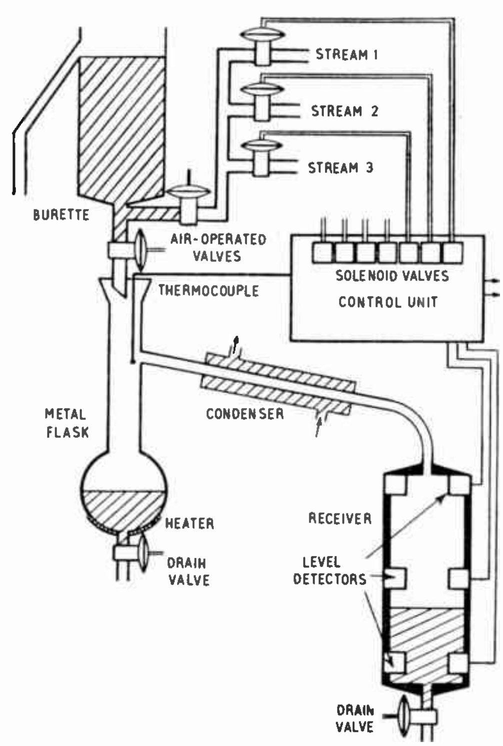
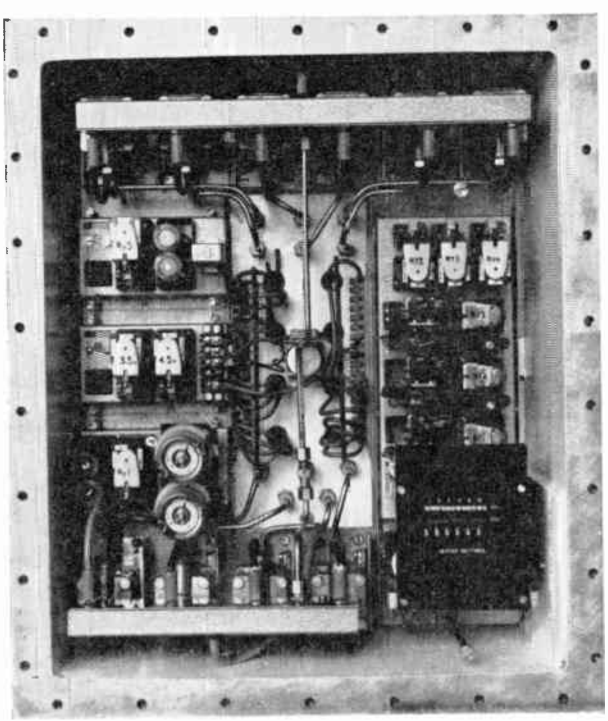


Fig. 2. Diagram of multi-stream distillation point monitor (BP-Hone).



safe, reliable, on-stream operation in a plant environment. A measured volume of the selected sample flows from a burette into a metal flask and is distilled into a receiver. (The receiver is fitted with pre-set photo-electric level detectors, which are of potted construction.) As the level of distillate rises in the receiver, these detectors actuate a printing or 'blipping' mechanism on the temperature recorder to indicate the temperatures (i.e. distillation points) corresponding to the selected percentages. A separate arrangement, associated with the novel heater design, is used to detect the final boiling point. At the end of each test, the heater switches off, the monitor drains and refills with the next stream in sequence, and the next test commences.

Distillation point monitors are now being manufactured in Great Britain and Fig. 3(*a*) shows one of these instruments and its sampling system installed at Kent Refinery. The control and programming unit, shown in Fig. 3(*b*) with the cover removed, is mounted on the monitor frame and contains the relays, peg board, and solenoid valves used to actuate the air-operated sample valves. The power supply unit and separate terminal box are mounted lower down the frame. All enclosures are air purged for safe operation and cleanliness.

The distillation monitor, in common with many other types of process analysers, operates on a cyclic basis and provides only an intermittent, instantaneous measurement of the desired property. To convert this into a continuous output signal suitable for automatic control applications requires the use of electronic 'peak-picking' or memory devices. This is relatively easy in the case of the distillation monitor, which can provide a memory actuating signal coincident in time with the desired value. For instruments such as the gas chromatograph described below, the memory device must detect and store the value corresponding to a peak maximum. Diode gating circuits charging a capacitor have been used for such applications.

3.1.3. Process chromatograph

Most petroleum products are very complex mixtures and their quality has to be defined by collective properties such as specific gravity, boiling range, viscosity, etc. However, for products such as bottled gas and petrochemical feedstocks, quality must often be specified in terms of chemical composition. For example, bottled gas may be a mixture of the hydrocarbon gases ethane, propane, isobutane and normal butane, and the proportions of these individual

Fig. 3. (*a*) Distillation point monitor and six-stream sampling system installed at Kent Refinery.

(*b*) Distillation monitor control unit with cover removed.

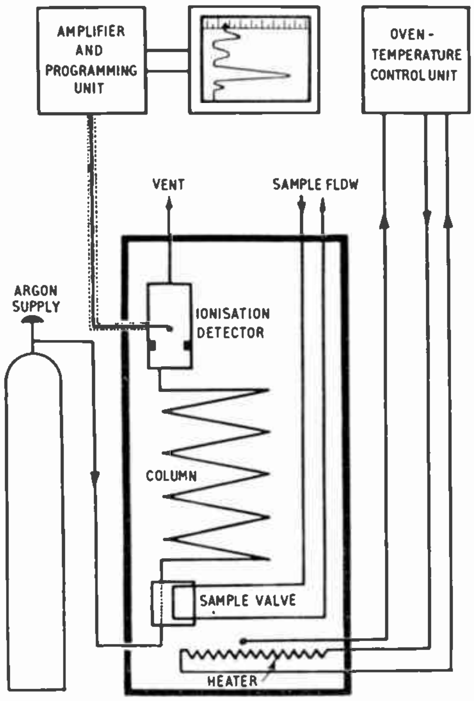


Fig. 4. Diagram of process chromatograph.

components can be measured by a relatively new technique known as gas chromatography. The technique is well suited for process monitoring and control applications, and Fig. 4 shows a simplified diagram of one of the process chromatographs used at Kent Refinery.

Briefly, a minute quantity (5 microlitres) of the gas to be analysed is injected automatically into a stream of argon flowing through a long, narrow tube (the column) packed with a granular adsorbent (e.g. alumina) which separates the components of the sample according to their boiling points. The separated components are eluted from the column as zones of hydrocarbon in the argon stream and are measured and recorded by the detector.

The most widely-used detector for process applications measures the thermal conductivity of the eluted gases, but this technique lacks the very high sensitivity required for some analyses. The Kent Refinery chromatograph uses an argon ionization detector, which measures the ionization current produced in the eluant gas stream by radiation from a radioactive strontium source. Currents are between 10^{-14} and 10^{-6} amperes, requiring a high impedance amplifier of good stability and linearity to produce a suitable output signal to the potentiometric recorder.

Figure 5 shows the prototype two-stream chromatograph installed on a plant. The three cylindrical steel casings are, from left to right, the analyser unit oven (containing sample valve, column and detector), the electrical unit (amplifier and programmer), and the oven temperature control and safety cut-out unit. The sampling system with stream selector valves and flowmeters is at the bottom left, and the flameproof isolating and air-purge pressure switches are on the right. The recorder is mounted in the control room and continuously records the composition of two gas streams as a series of peaks, the heights of which are proportional to component percentages.

3.1.4. Colour monitor

The final processing stage in the manufacture of lubricating oil is a treatment with activated clay to improve the colour and appearance of the product. A continuous, accurate measurement of product colour makes possible a substantial reduction in operating costs.

Colour is a subjective property and therefore difficult to measure absolutely, but on this particular process application colour can be correlated with the absorption of light of a selected frequency band.

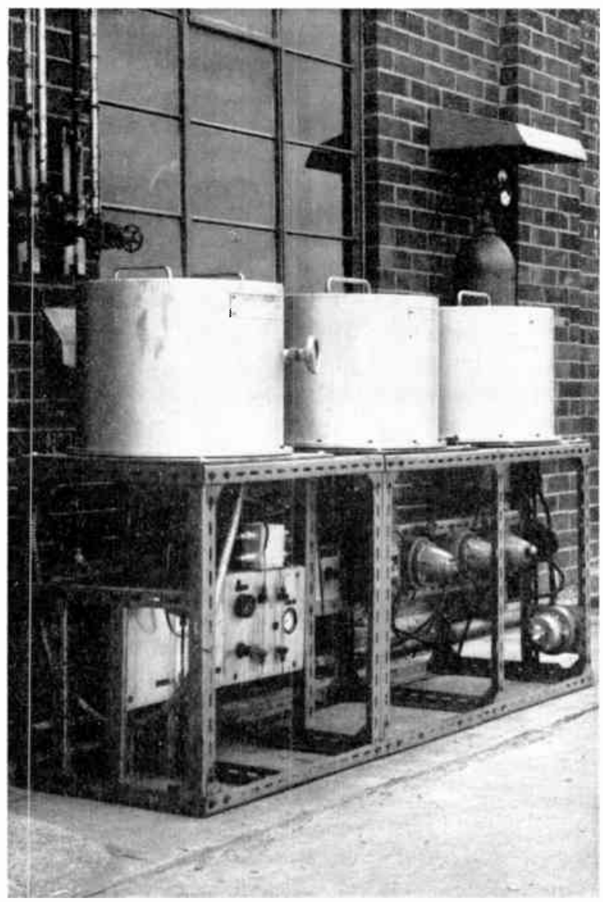


Fig. 5. Prototype BP process chromatograph installed at Kent Refinery.

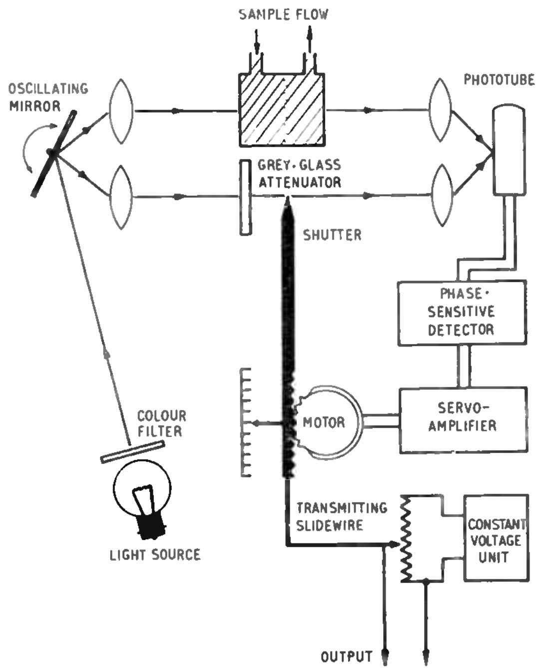


Fig. 6. Diagram of double-beam colour monitor (Sigrist and Weiss).

The principle of operation of the industrial null-balance absorptiometer used at Kent Refinery is shown in Fig. 6. White light from a tungsten filament bulb passes through a narrow band-pass filter and is chopped between the measuring and reference paths by a mirror oscillating at 550 c/s. Any difference between the absorption of the grey-glass standard and the continuously flowing sample produces an output from a phase-sensitive rectifier coupled to the photo-tube detector. This output operates the servo-driven shutter until balance is obtained. The shutter is mechanically coupled to a local indicator and transmitting slidewire. The instrument is at least twenty times as sensitive as the standard test for colour, and has proved to be very stable and reliable in operation.

3.1.5. Dielectric constant monitors

A range of quality measuring instruments which are purely electronic in operation are those based on dielectric constant. This property of oils and solvents can be correlated with other properties such as water content, the proportion of aromatic hydrocarbons, the concentration of metallic additives, etc. Dielectric constant can be measured continuously and accurately, using an a.c. bridge circuit to measure the capacitance of a cell through which the sample is flowing.

The method is used to detect water in the crude oil feed to distillation units, and thus reduce the loss in throughput which may occur as a result of process upsets caused by water. Similar instrumentation may improve efficiency of reforming processes producing components of high octane number for motor and aviation spirits. The difference in octane number between feed and product is related to the change in aromatic hydrocarbon content, and this can be correlated with the change in dielectric constant.

3.2. *Future Developments in Analytical Instrumentation*

The primary function of the process analysers described above is to provide a continuous record of product quality and facilitate process control at or near the specification level. However, it is the quality of the total batch in the storage tank which is important, and an off-specification product can be accepted into storage provided it is subsequently corrected during the filling period of the tank.

There is obvious scope here for automatic devices to measure the cumulative deviation of quality from the specification value, or to compute the weighted average quality of the tank contents. Process monitors, operating in conjunction with flow meters and simple computing devices, might be used to provide a more accurate analysis of tank contents than is obtained by present tank sampling methods.

An alternative approach would be to develop analytical devices to measure product quality directly in storage vessels. Such devices might reduce considerably the amount of laboratory effort at present devoted to testing samples drawn from every incoming and outgoing cargo, and from each of perhaps several hundred storage tanks.

In future refineries, one can envisage quality sensing probes installed in each tank, transmitting data on selected physical properties to a central data processing unit. This would be a logical extension of the electronic systems at present used for remote transmission of tank levels and temperatures.

4. Quantity Measurement and Blending

Accurate measurements of quantity are essential to keep account of the continual movements of the immense quantities of different types of products in a refinery's storage tanks. The most widespread method at present employed to measure quantity is tank level gauging, and volumes entering or leaving a tank are calculated from level changes.

Electronic equipment is used extensively to measure levels and temperatures, and to transmit the information in analogue or digital form to a central receiving station. In some instruments, level is sensed by a float or displacer suspended on a wire, but as these devices are affected by the nature and density of the liquid, contactless level sensors have been developed using radio-frequency or capacitance probes. Most systems are servo-operated and have an accuracy of better than 0·1 in over a span of up to 60 feet.

Most finished oil products are prepared by blending the basic components produced by the various refinery processes. The most widely used technique is batch blending, in which the required quantities of each component are pumped into a storage tank and circulated to ensure a homogeneous blend.

Level gauging and batch blending both require the use of storage tanks, which are very expensive to install and to maintain. Therefore, continuous metering and in-line blending techniques are being introduced, to use existing tankage more effectively and reduce capital expenditure on tankage in new refineries.

At the present time, the two types of flow meters most commonly used on these applications are positive displacement (p.d.) meters and turbine meters. With the incorporation of photo-electric or magnetic pick-off devices, both types can provide an electrical pulse output proportional to flow rate. Meters of this type can be used in automatic in-line blending systems where the pulse frequency from each component meter is compared with a pre-set frequency, and any difference is converted into a control signal to adjust the blend proportions.[7]

Line blending by fixed volumetric proportions requires the quality of each component to be known and constant, if the blend is to be produced economically with minimum 'give away'. In circumstances where these requirements cannot be met, a continuous analyser may be used to monitor the quality of the blend and to adjust the blend proportions automatically to control product quality at the desired level.

Flow meters with electronic digital output are well suited for applications involving remote transmission, temperature or density correction, data processing and computer control.

5. Plant Inspection and Maintenance

Refinery plants are required to operate continuously for as long as two years between major overhauls. Plant shut-downs represent a serious loss of revenue and productivity, and must be kept to the absolute minimum. Modern instrumental techniques are being used increasingly for plant inspection and maintenance, and the present trend towards methods for inspecting a plant whilst still on-stream could significantly improve productivity.

Established refinery inspection techniques like ultrasonic thickness gauging are being supplemented by radiographic and magnetic measurements. Closed-circuit television could be used to inspect the inside of pipes and vessels which might otherwise be relatively inaccessible. Ignition analysis can improve the setting-up and performance of reciprocating gas engines, and vibration analysers are useful on high speed rotary pumps and turbines.

Major overhauls require careful planning if they are to be completed in the shortest possible time, therefore computers are being used to program the sequence of operations, to make the most effective use of labour and machinery.

6. Conclusions

From the foregoing examples it will be evident that electronic equipment and techniques are being applied increasingly in every phase of oil refining. It cannot be claimed that the applications are unique, or indeed fundamental. Nevertheless, electronics has become a valuable and indispensable tool in the search for higher productivity.

7. Acknowledgment

The author wishes to thank the Chairman and Directors of The British Petroleum Company Limited for permission to publish this paper.

8. References

1. W. T. Lee, "The economic justification of on line computers for process control". Paper presented at Brit.I.R.E. Convention on "Electronics and Productivity", Southampton, 1963.
2. O. L. Farrar, "Process instrumentation and control, Part 3", *Oil & Gas J.*, p. 84, 22nd October 1962.
3. G. C. Eltenton, "Some instruments for quality control in petroleum refineries", *J. Appl. Chem.*, p. 246, 4th May, 1954.
4. W. H. Topham, "A rapid method of distillation for the routine quality control of refinery products", *J. Inst. Petroleum*, **45**, No. 431, p. 335, November 1959.
5. C. M. Tyler, "Process analysers for control", *Chemical Engineering Progress*, **58**, No. 9, p. 51, September 1962.
6. "Distillation point monitor", *Chemical Processing*, p. 12, November 1962.
7. K. Brooks and H. C. Bozeman, "In-line gasoline blending on the move", *Oil & Gas J.*, p. 83, 8th October, 1962.

Manuscript first received by the Institution on 27th February 1963 and in final form on 9th September 1963. (Paper No. 862).

DISCUSSION

Under the Chairmanship of Mr. M. James

Mr. J. F. Roth: The reliability and standards of construction of electronic equipment used for industrial applications must be of the highest order to ensure adequate performance. Unfortunately, the user so often finds that the equipment supplied is inadequate for the application for which it is to be used. This all too often stems from the fact that too many designers and engineers have little idea of the environmental conditions that will be met by the equipment they have designed. A prime requirement is that they get out of their comfortable offices and into the field so that they can experience for themselves the conditions met in practice. For this, a visit of more than an hour or two is necessary. After working, say, the 10 p.m. to 6 a.m. shift for a couple of nights their ideas will be rather different from those they were previously putting forward!

Seeing that breakdown of equipment provides major problems when trying to maintain production levels, a significant contribution to improving productivity can be made in this direction. This correction is, after all, aimed at methods for improving productivity, and one very significant—if not very startling—contribution would be made if all those responsible for the design of industrial electronic equipment would resolve to make a determined effort to give adequate attention to this aspect of their work.

The Author (*in reply*): I agree with Mr. Roth that the performance and reliability of industrial equipment would almost certainly improve if more manufacturers were better acquainted with the user's problems and the process environment. Instrument manufacturers should supply a complete application service to their customers, and not be content with selling only hardware.

In our experience, manufacturers welcome every opportunity to discuss with users the design, performance and application of their equipment, to the mutual benefit of both parties.

Mr. R. L. Duthie: If the petroleum industry is dissatisfied with the electronic instruments it can obtain, could Mr. Topham and his colleagues be induced to prepare a specification perhaps on the lines of that presented by Shaw† on steel works equipment at last year's Symposium on Industrial Electronics?

The Author (*in reply*): The continuous analytical instruments described in the paper normally employ conventional, well-proven electronic circuits and techniques. In my experience the electrical equipment used is quite reliable, but the overall design and performance of the instruments often leaves much to be desired. The situation is gradually improving with increasing liaison and co-operation between manufacturers and users, in this relatively new field of instrumentation.

I would agree with Mr. Duthie on the need for a code of practice to cover the particular requirements of instrumentation for the petroleum and allied industries. Electrical safety is of the utmost importance in refineries, and the new draft code shortly to be published by the Institute of Petroleum should be of considerable value in this respect.

Mr. K. A. MacKenzie: Has the author experienced any difficulties in transmission of d.c. signals from the electronic control instruments on the plant back over relatively long lines to the central control room?

The Author (*in reply*): We have experienced no difficulty with the transmission of d.c. millivolt or current signals over distances of several hundred feet, using either unscreened cable (e.g. thermocouple lead wire) or screened cable (e.g. copper sheathed mineral insulated power cable). We have no experience at Kent Refinery of using electronic process control instruments transmitting over distances of several miles, as would appear to be possible.

† D. Shaw, "A specification for electronic equipment for use in heavy industry", *J. Brit.I.R.E.*, **24**, No. 2, p. 133, August 1962.

U.D.C. 621.396.96: 681.32

Digital Automatic Radar Data Extraction Equipment

By

J. V. HUBBARD, B.Sc.

(*Graduate*)†

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

Summary: This paper describes a method of extracting automatically the range and bearing of point source echoes in the video signals from a two-dimensional radar (i.e. a radar scanning in range and bearing, or elevation, but not both), employing a digital ferrite core store for the necessary video storage. Attention is drawn to the flexibility of such a storage medium, together with a brief description of this type of store.

A method by which the video signals can be quantized in range, bearing and amplitude, to render them suitable for digital storage is then described, together with a method of video condensation, to economize on the amount of storage capacity required. Consideration is then given to the criteria for the recognition of echoes in the stored video and for the determination of their centre of symmetry, together with examples of the type of circuit which will meet these criteria.

Finally the precautions necessary to avoid the generation of false alarms due to meteorological clutter are examined and the method by which the range and bearing (or elevation) of the echoes recognized, may be generated in a suitable form for transmission to a radar data handling computer.

1. Introduction

In the paper by J. C. Plowman‡ it is shown that, in order to detect automatically the presence of echoes in radar video, it is necessary to store the video signals for a period corresponding to the time taken for the aerial to traverse one aerial beamwidth. His paper then considers two possible types of video storage, namely storage tubes and ultrasonic quartz-delay lines. However it will have been observed that the quartz-delay-line system has to be tailored to the radar, i.e. the delay of each line must be very accurately matched to the radar pulse repetition frequency and there must be the same number of delay lines as there are pulses per aerial beamwidth. Thus a major re-design of the system is necessary to meet the parameters of each new radar, also if the radar employs many pulses per beamwidth the required number of delay lines would make the equipment prohibitively bulky and expensive.

To overcome these difficulties the possibilities of using a digital ferrite-core store, similar to those used as data stores in digital computers, was considered for video storage. Such a store has the advantage that it is aperiodic, that is, that information can be read out of the store at any time after it has been injected, as determined by the addressing program. This type of store can therefore be made to cope with widely varying radar parameters, simply by changing the program. This paper describes how such a store can be used for video storage and some of the types of echo-recognition circuit which may be used with it.

2. The Ferrite-core Store

As this type of store is now widely used in digital computers, only a brief resumé of its principle of operation will be given. Basically, the storage element consists of a ferrite ring, approximately 1 mm dia. made of a magnesium manganese ferrite which has a substantially square hysteresis loop, i.e. the properties of a permanent magnet. Such an element can be used as a memory by its ability to retain either a clockwise or counter-clockwise field, which may be used to represent a 'one' or a 'zero' respectively. The core is switched to the required state by passing the appropriate magnetizing current through the ring. To avoid the large number of drive circuits necessary to supply these switching currents when many cores are used, the cores are arranged into the form of a matrix with switching wires running vertically and horizontally through all cores, as shown in Fig. 1. Thus any individual core may be switched by applying half the necessary switching current to both the

† Admiralty Surface Weapons Establishment, Hampshire.

‡ J. C. Plowman, "Automatic radar data extraction by storage tube and delay line techniques". *The Radio and Electronic Engineer* (*J.Brit.I.R.E.*), **26**, No. 4, pp. 317–26, October 1963.

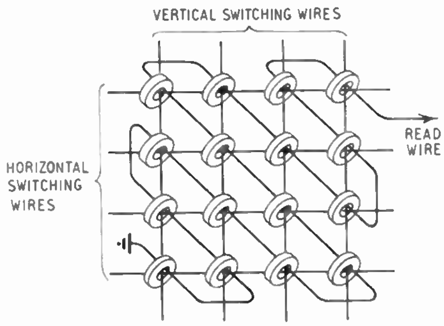


Fig. 1. Ferrite core store.

vertical and horizontal wires which intersect at the required core. (Due to the square hysteresis loop property of the ferrite, there will be negligible change in the field of those cores which are only linked with one of the energized wires.) This arrangement reduces the number of drive circuits required to twice the square root of the number of cores in the matrix.

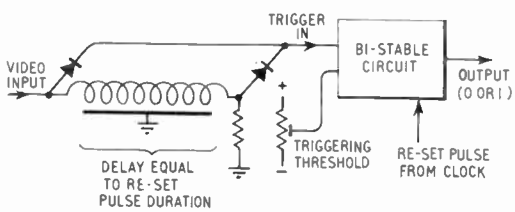


Fig. 2. Video encoder.

The information thus stored in a core is read out by applying the required drive currents to switch it to 'zero'. If the core was in the 'one' condition, there will be a reversal of flux in the core which will generate a voltage pulse in a 'read' wire which is also linked with the core. If the core was already in the zero condition there will be no change of flux and no voltage pulse. As a voltage pulse can only be obtained from the core which is switched, a single read wire may be used which links all cores in the matrix.

It will be noticed that reading information out of the store will erase it from the store, hence if it is required to refer to information in the store without losing it, it must be re-written. The time required to carry out this read and write process is known as the store cycle time and is usually within the range 1 to 10 microseconds.

As it is often desirable to read the information from a number of cores simultaneously, a number of matrixes are built up into a stack and the switching wires to each matrix are connected in parallel. Thus by energizing a particular pair of switching wires, the condition of the appropriate core in each matrix can be read from the matrix read wires, simultaneously. Those cores which can be read in parallel are referred to as a store word, thus the number of elements or 'bits' which form a word is determined by the number of matrices in the stack and the number of words is determined by the number of cores in each matrix. Each word of the store will have an 'address' determined by the combination of switching wires to select the appropriate cores and, as it is desirable to generate the address in binary form, the number of vertical and horizontal switching wires will each be a power of two, usually the same. Hence the number store of words will be an even power of two, 256, 1024 and 4096 being common values. There is no restriction on the number of bits per word other than the complexity and cost as the associated read and write circuitry is directly proportional to the number of bits per word, whereas the addressing circuitry only increases as the square root of the number of words. Thus it is usual to restrict the number of bits per word to 30 or less.

2.1. *Video Quantization*

As a digital store can hold only 'ones' and 'zeros' it is necessary to reduce the video signals to this form before they can be stored. Hence the video must first be quantized into increments of amplitude and range, the bearing having already been quantized by the radar pulse repetition intervals. Video amplitude can be quantized by comparing the signal with a threshold and generating a 'one' whenever the threshold is exceeded. Multi-level amplitude quantization can be achieved by employing a series of such thresholds and encoding a count of the number of thresholds which have been crossed, in binary form.

Range quantization requires each range sweep period to be divided into a number of discreet increments. In order not to degrade the inherent range resolution of the radar these increments should be made equal to the radar pulse length. In practice range and amplitude quantization can be carried out together by feeding the video signal into a bistable circuit with the required triggering threshold and re-setting the bistable circuit, if triggered, at the end of each range increment. The re-set pulses are generated by a 'clock' oscillator with a pulse repetition interval equal to the required range increment size. This clock oscillator is re-started in phase with the radar trigger at the commencement of each range scan. To avoid missing a video peak which might have triggered the bistable circuit during the re-set pulse period, the signal applied to the bistable circuit should be the incoming video or the incoming video delayed by the re-set period, whichever is the greater, as shown in Fig. 2.

2.2. *Video Storage*

Having processed the video signal into a series of zeros and ones so that it can be accepted by the store, it is necessary to consider how this information should be arranged in the store to facilitate its examination by the echo recognition circuits. It will be seen later when the echo recognition circuits are discussed that they must correlate the video levels across the radar aerial beamwidth at a given range. As all the bits of a given word can be made available simultaneously, it is desirable that the bits of each word should represent the signal levels across the beamwidth in order that they may be correlated, using one word per range increment. Each level stored in the word may consist of one or more bits depending on whether single- or multi-level quantization is used. This configuration is also desirable economically, as for most radars there are considerably more range increments (i.e. pulse lengths per range sweep period) than pulses per beamwidth. The store will thus hold the signal levels from a sector of angular dimension equal to the aerial beamwidth and radius equal to the maximum range over which auto-detection is required.

Assuming the store has already been filled, it is necessary to keep it up-dated so that it always holds the signal levels received over the previous beamwidth of aerial rotation. This is achieved by shifting all the levels, already stored in the word, along by one and inserting the new level just received into the now vacant first level position, the level previously held in the far end of the word being discarded, see Fig. 3. This process is repeated for each store word in

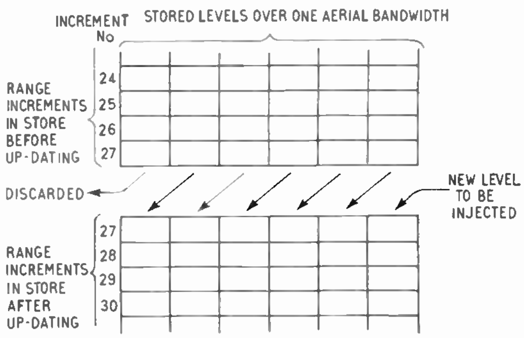


Fig. 3. Up-dating of moving window words.

turn until all range increments have been up-dated, starting again at the first range increment with the next radar trigger pulse. The store can thus be regarded as a 'moving window' traversing in azimuth in synchronism with the aerial.

As one store word must be addressed by each range increment, it is essential that the store cycle time is not greater than the radar pulse length, otherwise there will be a loss of range resolution and detection sensitivity.

2.3. *Video Condensation*

Even though the video storage has been arranged so that the economically desirable arrangement of more words than bits per word has been achieved, the number of bits per word may still be economically impracticable if the radar employs many pulses per beamwidth or if multi-level video quantization, demanding several bits per level, is required. Under these conditions, the word length can be reduced by reducing the number of levels stored over the beamwidth. Although this could be achieved, simply by counting down the radar pulse repetition frequency and only storing the video from say, every third range sweep, thereby reducing the word length by a factor of three, this would give rise to a loss of sensitivity due to the fact that available radar data was being ignored. Such a loss can be avoided, in the example quoted, by storing the mean of the levels received over groups of three pulse repetition intervals. This process is referred to as video condensation and the number of range sweeps to be averaged is known as the condensation factor C.

The average level from each range increment is obtained in the normal way by adding the encoded levels received on C successive range sweeps, dividing by C and injecting the quotient into the moving window word. In order to hold the sum for each increment between range sweeps additional storage is required. This storage can use either additional bits in the moving window words or a separate group of addresses in the same store. In the latter case there would have to be twice as many store words as range increments although, as previously stated, it is often more economical to increase the number of words rather than the word length.

In either case, the video sum is obtained by reading each condensation store word in turn, digitally adding the new encoded video level to it and re-storing the result as shown diagramatically in Fig. 4. To avoid the increase in range increment size, which would otherwise be necessary to allow for the operating time of the adding circuits, a running address can be employed in which the up-dated sum is re-stored in the next range increment word as that value is read out for up-dating. This allows a whole store cycle period for the addition without affecting the range increment size. The only penalty with this system is the necessity to provide suitable logic to ensure that the store address, corresponding to zero range at which to start the up-dating process, steps on by one with each radar trigger.

When the condensation factor C is a power of two, the necessary division can be achieved by a shift to the right of $\log_2 C$ bits, discarding the least significant digits. If C is not a power of two, an approximation to true division (which must be carried out within the store cycle time) is obtained by applying two different shifts and adding or subtracting the numbers thus formed. This is equivalent to multiplying by the expression $\frac{1}{2^x} \pm \frac{1}{2^y}$, the values of x, y and the sign being chosen to give the best approximation to $1/C$.

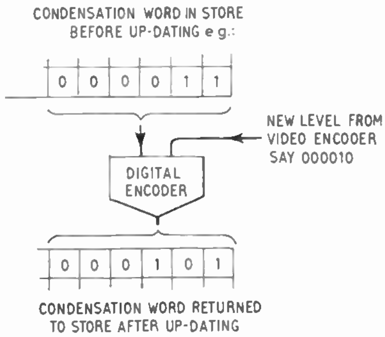


Fig. 4. Up-dating of condensation words.

If the condensation sum is held in the same word as the moving window, the latter can be up-dated by arranging for the shift required for division to move the result into the appropriate bits of the moving window part of the word. This shift is only applied after every C radar pulse repetition intervals, i.e. when a new condensation sum has been formed.

If the condensation sum is held in a separate word from the moving window, it will be necessary to effect a transfer from the condensation word to the moving window word. This requires the two words to be addressed alternately (as it is not possible to address two words simultaneously) requiring two store cycles per range increment. If the transfers were carried out at the same time as the video is being received, known as live time, this would restrict the minimum radar pulse length to two store cycle periods. However, this limitation can be overcome by carrying out the transfers during dead time i.e. the time between the receipt of the video level from the last range increment and the arrival of the next radar sync pulse, corresponding to the fly-back time of a visual display. An excessively long dead time can be avoided by virtue of the fact that each moving window increment has only to be up-dated every C pulse repetition intervals, consequently it is only necessary to up-date $1/C$ of the moving window words in any one dead time.

2.3.1. Choice of condensation factor

The choice of condensation factor will depend on two requirements, firstly on the desirability of using a large condensation factor to economize on the word length to hold the moving window, and secondly on the requirement for bearing accuracy. Regardless of the type of echo discriminator used on the data stored in the moving window, it cannot select the echo centre to better than the nearest sample stored therein i.e. to better than plus or minus half the angle corresponding to the number of radar pulse repetition intervals condensed. Consequently if, for example, a bearing accuracy of plus or minus a tenth of a nominal beamwidth is required, there must not be less than five condensation samples per nominal beamwidth.

Thus, once the necessary circuits have been introduced to enable the video to be condensed, the equipment becomes very versatile as it is only necessary to change the condensation factor to enable it to operate on the signals from radars employing widely different numbers of pulses per beamwidth.

3. Echo Discriminators

An echo discriminator is required to fulfil two functions, namely to give an indication when the level pattern stored in the moving window resembles that which is expected from a point target and secondly to give this indication only when the target signals are symmetrically placed in the moving window word.

As has been previously stated, the criterion by which a radar echo from a point target is recognized is that it will produce an increase in the mean signal level over one pulse length at a constant range (corresponding to that of the target) covering a bearing of from one to possibly two and a half nominal aerial beamwidths, depending upon echo amplitude, aerial polar diagram, etc. On the other hand, signals other than from point targets will in general have either or both the range or bearing dimension increased. Consequently, the echo discriminator must examine the stored levels, range increment by range increment, to determine whether the mean of the levels over the aerial beamwidth in any increment is sufficiently above the mean due to the receiver noise to indicate the presence of a target. In addition, it must check that the area over which these increased levels have been received are within the limits expected and it should delay the indication of the presence of a target until the pattern of levels is symmetrical about the centre of the moving window.

3.1. *One-bit Echo Discriminators*

The simplest type of echo discriminator is one which operates on the stored signals from a single threshold video encoder. If this threshold is set so that the probability of generating a 'one' in the moving window due to receiver noise is low, the probability of obtaining a group of ones in any one moving window word will be very low. On the other hand, the presence of even a small echo in the video will appreciably increase the probability of the signal crossing the threshold, giving a high probability of obtaining a group of ones in the corresponding moving window word. Consequently the primary task of the echo discriminator is the ability to sense the presence of a series of ones in a moving window word. This task can be achieved by feeding each moving window word in turn on to a register (as it is addressed for updating) and generating an analogue sum of the number of ones held on the register by a resistance adding circuit as shown in Fig. 5. The output of this circuit can then be fed to a comparator which will indicate a detection whenever the sum of the ones exceeds a pre-set threshold.

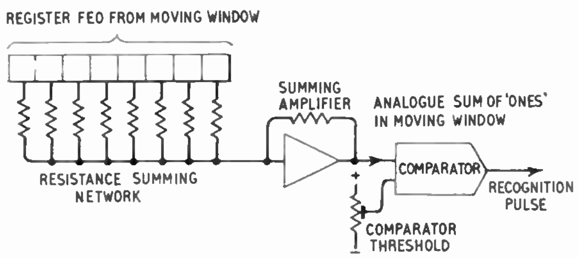


Fig. 5. Basic one-bit echo recognition circuit.

Although this simple circuit can indicate the presence of an echo, it is unable to sense when it is centred in the moving window or whether its dimensions are within the expected limits. A maximum limit to the echo bearing can be set by arranging for the moving window word to cover a bearing greater than that expected from the largest echo of interest and inhibiting the output whenever a 'one' is present on either of the end bits of the register, as shown in Fig. 6.

The sensitivity and bearing accuracy can be optimized by symmetrically weighting the summing resistors to give greater weight to 'ones' in the centre of the moving window word. Thus any pattern of 'ones' shifting through the moving window word will give maximum output when the pattern is symmetrically placed about the centre of the moving window word. The point at which the maximum output is obtained can be determined by employing a second identical weighted summation circuit connected one bit nearer the input end of the moving window word. This circuit, will then, in effect, have a 'pre-view' of the weighted sum which will be generated in the first circuit after the next moving window shift, see Fig. 7.

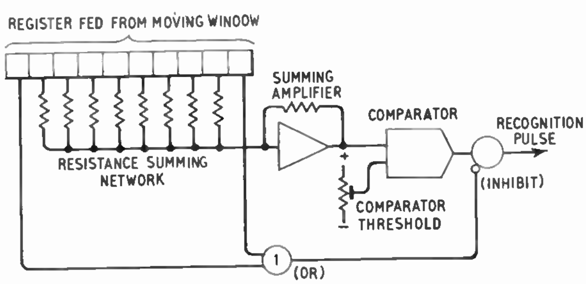


Fig. 6. Recognition circuit with wide echo inhibit.

If the indication of the detection is delayed until the first occasion on which the weighted sum from the 'pre-view' circuit is less than that from the first circuit, this will indicate that after the next moving window shift, the output of the first circuit will have just started to fall off, hence it must now be at its maximum.

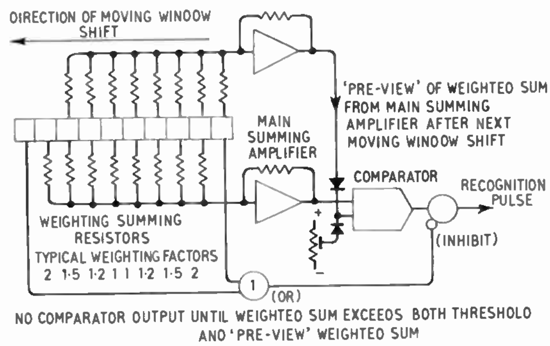


Fig. 7. Recognition circuit with maximum symmetry criterion.

This can be achieved by inhibiting an indication of detection from the first weighted sum circuit until its output exceeds either the threshold or the 'pre-view' weighted sum, whichever is the greater, as shown in Fig. 7.

The range dimension of the echo can be checked by a logical circuit which will reject recognition pulses which occur in three or more successive range increments, as not emanating from a point target. It is necessary for this circuit to accept recognition pulses from two successive range increments as these may be due to a point target echo which happens to bridge the interval between two range quanta.

When two such recognitions occur, only the first will be passed out as a target position.

3.2. *Digital One-bit Echo Discriminator*

An alternative approach to the one-bit echo discriminator is to operate logically on the bits pattern contained in the moving window word according to the following rules:

(*a*) The centre three bits of the word must be 'ones'.

(*b*) The end bits of the word must be 'zeros'.

(*c*) The number of ones in the remaining bits to the left of the centre must be equal or one greater than the number of 'ones' in the remaining bits on the right.

Rule (*a*) is a form of weighting taken to the extreme in which only the centre three bits count at all. The value of three has been chosen on the assumption that the video encoder would be set to give approximately a 0·01 probability of generating a 'one' in the moving window due to noise. Hence the probability of noise generating three 'ones' in the selected bits is approximately $(0{\cdot}01)^3$ or 1 in 10^6 which is an acceptable probability of a false alarm.

Rule (*b*) is the maximum echo width criterion as used in the analogue circuit.

Rule (*c*) ensures that a detection will only be made when the 'one' pattern is symmetrical about the centre of the moving window word. The tolerance of one on the symmetry criterion is essential in order to accept patterns consisting of an even number of 'ones' which cannot achieve exact symmetry about the three compulsory bits.

3.3. *Disadvantages of One-bit Echo Discriminators*

The first disadvantage of a one-bit echo discriminator is the large tolerance necessary on the permissable echo width if a large range of target amplitudes is to be accepted. Although the echo amplitude variation due to target range can be eliminated by applying accurate swept gain (s.t.c.), the echo discriminator has still to cope with the range of target echoing areas of interest. The effect of the large tolerance in echo width is to permit detection of small non-point target echoes provided the width of the return does not exceed the expected width of the strongest point target of interest.

The second disadvantage is that in order to be able to detect weak signals the video encoder threshold must be situated as low as possible in the receiver noise commensurate with an acceptable false alarm rate. Consequently the patterns of 'ones' in the moving window even from strong echoes is liable to be disturbed by noise at the edges which will prevent the achievement of the maximum possible bearing accuracy.

Both these disadvantages can be overcome by employing multi-level video encoding together with a suitable echo discriminator.

3.4. *Multi-bit Echo Discriminators*

The first step on from one-bit working is two bits employing three video encoder thresholds, the output encoded in the form:

00 if no threshold is crossed

01 if the first threshold is crossed

10 if the second threshold is crossed

11 if the third threshold is crossed.

The simplest form of echo discriminator to avoid the disadvantages enumerated in Section 3.3 above is a set of three one-bit discriminators as described in Sections 3.1 and 3.2 operating at each of the three threshold levels, the outputs being OR'd together. This can be achieved by connecting the first to the least significant level digit, the second to the most significant level digit and the third to the two digits AND'd together. Thus each discriminator will have to cope only with a third of the echo amplitude range covered by a single one-bit discriminator, enabling a smaller tolerance to be employed on the acceptable echo width at each level. The bearing accuracy will also attain its maximum once the echo amplitude is sufficient to permit detection at the second level as this level is well clear of noise and the pattern of ones will be dependent entirely on the echo.

It will be noted that the above echo recognition criteria in effect are crudely testing whether the received pulse amplitude pattern corresponds to the two way beam shape pattern of the aerial, i.e. for a given echo amplitude it checks whether the width of the pattern is within certain limits determined from the known aerial polar diagram.

In fact the logical conclusion to multi-bit working is to encode and store the video amplitude in a sufficient number of bits to enable a beam shape discriminator of the type described by Plowman,† to be applied to the stored video, after reconversion to analogue form. In order to benefit from the greater discrimination of this type of echo recognition circuit against signals from other than point targets, it is necessary to encode the video to a minimum of four bit accuracy. The increase in word length to store the levels to this accuracy will be partially compensated for by the fact that it is only necessary to make a beam shape comparison over from one to one and a half nominal beamwidths, whereas a one-bit discriminator must operate on the stored pattern over an arc greater than that subtended by the largest signal of interest which may be from two to two and a half nominal beamwidths.

4. Clutter Suppression

So far it has been assumed that the background against which detections have to be made is the thermal noise generated in the receiver, whereas in fact the background may consist of meteorological or land clutter in addition to the receiver noise. Consequently an automatic detection equipment must be designed so that it will not generate false alarms in clutter areas but will continue to detect any echoes of sufficient amplitude to show above the clutter.

As clutter signals are produced by the addition of returns, which will be in random phase, from innumerable small reflecting particles, the resultant amplitude distribution will closely approach the Gaussian distribution of receiver noise. When such signals have been detected by a logarithmic receiver the a.c. deviation component is independent of the signal

† J. C. Plowman, *loc. cit.*

amplitude, only the d.c. component increases with increasing signal power. Hence provided the video encoder threshold can be related to the mean d.c. level of the background, the echo recognition requirements will be the same for a clutter background as for noise.

Although the d.c. component of the signal can be removed by a simple *R–C* 'differentiating' circuit giving good clutter suppression, there will be an appreciable loss of sensitivity to genuine targets. This loss of sensitivity which is of the order of 5 dB is due to the reverse polarity 'overshoots' obtained when pulses are applied to an *R–C* differentiating circuit. This causes the large negative going spikes, which are characteristic of logarithmic noise, to add positive going overshoots to the normal positive excursions of the noise, which will tend to generate false alarms. The increase in the video encoder threshold necessary to bring the false alarm rate back to the permissible level, produces the loss of sensitivity observed.

This loss can be avoided by employing a delay line circuit to subtract the d.c. component of the signal. In this circuit the video is applied to a terminated delay line of the order of ten radar pulse lengths in length (Fig. 8). The mean d.c. level of the background is obtained by a resistance adding circuit connected to a series of taps down the line, excluding the centre tap. This d.c. level is subtracted from the signal sample taken from the centre tap in a difference amplifier. With this circuit, no measurable loss of sensitivity was observed although it tended to allow the generation of false alarms on the leading and trailing edges of banks of clutter.

This was due to the fact that at the instant at which the leading edge of a bank of clutter reaches the signal sample tap, the clutter signals will be present on half the line only, consequently only half the mean d.c. level of the clutter signal will be subtracted from the signal sample. This failing can be overcome by modifying the circuit as shown in Fig. 9 so that the d.c. level to be subtracted from the signal is the sum of the levels over the first or last half of the line, whichever is the greater.

The effectiveness of this circuit can be seen from the photograph in Fig. 10 which shows a typical p.p.i. radar display with rain clutter present. Superimposed on the radar picture are the auto-detection output pulses which appear as short radial 'ticks' adjacent to the echoes. From this picture it can be seen that although the aircraft echoes are being detected, there are no false alarms on the clutter. It will however be noted that there are a large number of detection 'ticks' on the land returns near the centre of the p.p.i. This is due to the fact that land clutter, in addition to generating a background of more or less Rayleigh amplitude distribution, will contain a number of point reflectors such as large buildings etc. which will produce echoes which are indistinguishable from those due to aircraft. It is therefore not possible to differentiate between such echoes in the automatic detection equipment, the only method of elimination being the application of some form of moving target indication (m.t.i.) to the radar.

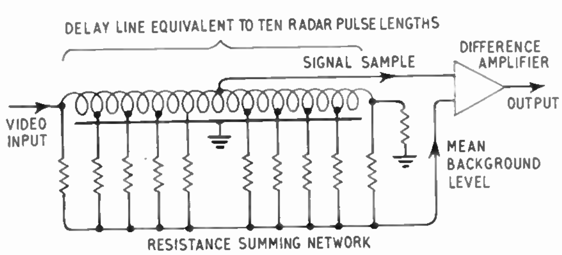


Fig. 8. Mean background level subtraction circuit.

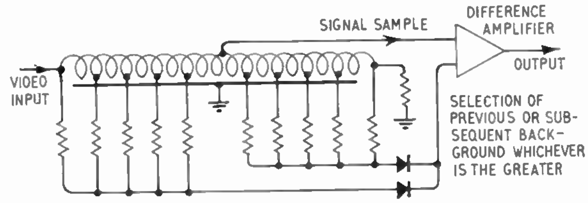


Fig. 9. Subtraction of previous or subsequent background whichever is the greater.

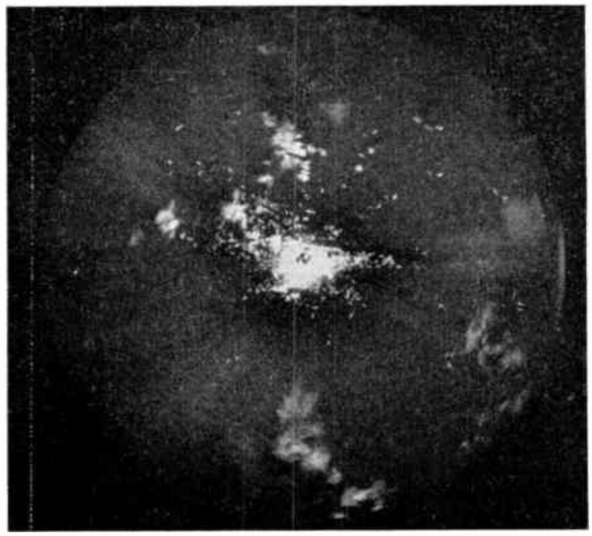


Fig. 10. Automatic radar detection in the presence of rain clutter. Targets detected automatically are indicated by the short radial 'ticks' adjacent to the echoes. It can be seen that no false alarms have been generated by the rain clutter.

5. Target Position Read-out

The echo recognition circuits described above will generate a recognition pulse whenever there is a pulse pattern, resembling that from a point target, centred in the moving window. Such a signal would not however be suitable for feeding a data-handling computer which requires target information in the form of range and bearing as binary words.

The range word can be generated by using the recognition pulse to staticize a binary count of the number of range increments out from the radar sync pulse. Similarly the angle at which the detection occurs can be read out by staticizing the output from an angular input, analogue/digital converter, rotating in synchronism with the radar aerial. As the signals from the whole of the target must be stored before an echo can be recognized, the echo recognition pulse will occur half a beamwidth late, hence it will be necessary to apply a bearing correction, either in the data handling computer or by suitably phasing the aerial rotation digitizer.

To permit read-out of target positions which may occur in rapid succession, it is necessary to read the range and bearing staticizers as soon as possible after they are set, so they are free to hold the next target position. As it is wasteful of computing time to interrupt a data handling computer at frequent and irregular intervals, it is desirable to employ a buffer store between the auto-detection equipment and the computer. This will receive and store the target position data from the auto-detection equipment as it is generated and, at suitable pre-determined intervals in the data handling program, transfer all new stored data to the computer.

6. Conclusions

Tests carried out on an experimental model of the digital radar data extraction equipment have shown that a system as described above is completely practical and has a detection sensitivity comparable with that of either a quartz delay line storage system or an alert visual operator. Range and bearing accuracies were good, accuracies of the order of plus or minus one range increment and plus or minus 10% of the nominal aerial beamwidth being achieved. The experimental model also proved the versatility of the digital approach and it was possible to use the equipment with radars employing widely different pulse lengths, pulse repetition frequencies and pulses per beamwidth, simply by adjusting the program to give the required range increment size and condensation factor. The only limitation is the minimum radar pulse length that can be used as determined by the store cycle time.

The experimental model also proved the great advantage of an automatic detection system, namely that the detection sensitivity is independent of time and target density; whereas a visual operator tires rapidly and becomes overloaded if he has to track many targets, with consequent loss of sensitivity and accuracy. On the other hand the target handling capacity of the auto-extraction equipment is limited only by the capacity of the associated buffer store and data handling computer.

7. Acknowledgments

This paper is published by permission of the Board of Admiralty. The author wishes to acknowledge the contributions of his colleagues at A.S.W.E. to this work.

8. Appendix: Terminology

Aerial Beamwidth. The arc of aerial rotation over which a useful signal may be expected from a point target. This is of the order of one to two nominal aerial beamwidths, depending upon the required range of echo signal to noise ratios, shape of the aerial polar diagram and the criteria by which the echo is to be recognized.

Nominal Aerial Beamwidth. The angle between the two points on the aerial polar diagram at which the aerial gain has fallen to 3 dB below maximum. In the case of radar operation where the aerial is used for both transmission and reception, the signal will be 6 dB below peak at these points.

Point Target. A target whose dimensions are small compared with a range difference corresponding to the radar pulse length and the linear length of the arc corresponding to one nominal beamwidth at the minimum useful range.

Manuscript received by the Institution on 1st April 1963. (Paper No. 863/RNA22).

(Discussion on this paper appears on opposite page)

DISCUSSION

Under the chairmanship of Mr. R. N. Lord

Mr. D. H. R. Archer: I may have misunderstood Mr. Hubbard, but I have the impression that even where several recognition thresholds are used there exists a finite probability that two adjacent targets separated by less than a beamwidth—may be rejected by the excessive width inhibiting circuits. If this is so, it represents a substantial operational hazard, since closing targets may be lost at a critical moment.

The Author (*in reply*): Mr. Archer is correct in assuming that there is a possibility of rejecting two targets within the same range increment when their bearing separation is of the order of one beamwidth. This is not however considered to be a more serious operational hazard than the fact that the two echoes would have merged into one on a visual display. It must be remembered that the data extraction equipment is intended to feed into a computer with the capability of rapidly and accurately extrapolating tracks, hence the ability to foresee a hazardous situation and advise avoiding action long before the target separation has been reduced to one beamwidth. This ability to extrapolate would also enable the computer to continue tracks through the crossing zone when targets cross at different heights (or otherwise without collision).

Mr. T. F. Spriggs: If I have understood these papers correctly they have described solutions to *all* of the present m.t.i. problems, namely

zero speed target (ground clutter, etc.)

low speed targets (rain, helicopters, etc.)

high speed targets

targets with zero radial speed but periphery speed (tangential tracks)

Can Mr. Hubbard confirm that we have now a full answer to the current m.t.i. system.

The Author (*in reply*): Although we would not be so bold as to say that we have the full answer to the current m.t.i. system, most of the target discrimination problems which can be solved by an m.t.i. system can also be solved by an auto-extraction equipment plus tracking computer combination. Firstly the auto-extraction equipment is capable of rejecting echoes emanating from reflecting objects whose dimensions are greater than the area of resolution of the radar, i.e. an area equivalent to the pulse length × the beamwidth. This enables it to reject rain clutter and the majority of land clutter. It is not, however, capable of rejecting reflections from prominent objects which stand out from the general land clutter, i.e. reflections from large buildings, pylons, etc. which produce echoes which are indistinguishable in form from aircraft echoes. Secondly, as the tracking computer is capable of determining the speed of all targets tracked, a minimum speed limit can be applied to the target markers which are presented for display. Such a speed limit would be independent of the direction of movement of the target. However it must be remembered that the computer must track a target in order to determine its speed and this could impose a heavy load on the computer if many targets outside the required speed range are present. On the other hand if the majority of the unwanted targets detected by the auto-extraction equipment are stationary, these can be rejected by storing their positions within the computer and ignoring future detections at these positions. Unfortunately, post-detector processing cannot provide velocity-selective sub-clutter visibility.

Brit.I.R.E. GRADUATESHIP EXAMINATION, MAY 1963—PASS LISTS

The following candidates who sat the May 1963 examination at centres outside Great Britain and Ireland succeeded in the sections indicated. The examination, which was conducted at 86 centres throughout the world, attracted entries from 592 candidates. Of these 202 sat the examination at centres in Great Britain and Ireland and 223 sat the examination at centres overseas. The names of successful candidates resident in Great Britain and Ireland were published in the November issue of the *Proceedings* of the Brit.I.R.E.

Section A	*Candidates appearing*	*Pass*	*Fail*	*Refer*
Great Britain	74	26	45	3
Overseas	103	22	72	9
Section B				
Great Britain	128	47	55	26
Overseas	120	24	73	23

OVERSEAS

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

BEN-YOSSEF, Michael (S), *Givatajin, Israel.*
BHANU-PRAKASH, B. V. (S), *Bangalore 11, S. India.*
CASSAR, Carmel Laurence (S), *Malta.*
CHAN, Fook Cheong (S), *Singapore 12.*
CHATURVEDI, Ram Nath (S), *Bombay 71, India.*
CHONG, Hoo Yuen (S), *Singapore 1.*
CHOPRA, Omprakash (S), *Delhi 6, India.*
DESAI, Ramesh Chandra (S), *Breman, W. Germany.*
FRISCH, Abraham (S), *Holon, Israel.*
HILLMAN, Peter (S), *B.F.P.O. 180.*
JOSHI, Ganesh Bahirao (S), *Bombay 16, India.*
KARKHANIS, Bhalchandra, *New Delhi, India.*
MACKINNON, Terence Charles Flynn (S), *B.F.P.O. 53, Cyprus.*
NARAYANA-RAO, N. (S), *Bangalore 4, India.*
NISSIM, Moshe (S), *Ramat Gan, Israel.*
PANG, Ee Ang (S), *Singapore 3.*
PARANN, Moshe, *Tel-Aviv, Israel.*
PATHMANATHAN, Somasundran, *Kuala Lumpur, Malaya.*
PUNIA, Atma-Singh (S), *Agra, India.*
RAJPUT, Dhir Singh (S), *Punjab, India.*
RAMAMURTHY, Aiyasamy (S), *Bangalore 6, India.*
RENDALL, Gavin Mowat, *Lagos, Nigeria.*
ROSSI, Francis (S), *Sliema.*
SAVVA, Costas, *Nicosia, Cyprus.*
SWAROOP, Jagdish, *Bombay 16, India.*

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

AHRON, Edelstein (S), *Jerusalem.*
ALTARATZ, Jacob (S), *Tel-Aviv, Israel.*
BALAKRISHNAN, K. (S), *Truhur, Kerala State, India.*
CLARKE, Rodney Dallas, *Victoria, Australia.*
DOWNEY, Arthur Charles, *Victoria, Australia.*
GAUNT, John Richard (S), *Fiji.*
GOPALKRISHNAN, Kollengode Sargameswara (S), *Bombay 71, India.*
HARIHARAN, Thiruvenkatan (S), *Mattancherry, Cochin 2.*
KRISHNASWAMY, Chakravarti (S), *Calcutta 26, India.*
PARULKAR, Subbash Ramrishna (S), *Cochin 4, India.*
PEARSON, Peter John, *Malta.*
RAO, Mysole Narayana Rama (S), *Bangalore 11, India.*
RILEY, Stanislaus John, *Fremantle, Australia.*
SAUND, Pritam Singh (S), *56 A.P.O. India.*
SEN, Naresh Chandra (S), *Nevallois, France.*
SREIDHARA-MOORTHY, N. S. (S), *Bangalore 19, India.*
STANISLAUS, Joseph (S), *Bangalore, India.*
SUNDARAM, Kalyana (S), *Mhar (M.P.), India.*
TENNENT, Donald Campbell, *Watsons Bay, N. South Wales, Australia.*
TOMLINSON, George Graham, *Lusaka, N. Rhodesia.*
TYROLER, Boas (S), *Kiryot, Ono, Israel.*

The question papers set in Section A and Parts 3 and 4 of Section B of the May 1963 Graduateship Examination have been published in the October and November 1963 issues of the *Proceedings* of the Brit.I.R.E., together with answers to numerical questions and examiners' comments. Part 5 of Section B will be published in the December *Proceedings.*

(S) denotes a Registered Student.

U.D.C. 621.396.96: 681.3

Design and Application of Computers for Radar Data Processing

By

R. A. BALLARD, B.Sc.†

AND

L. MOORE, B.Sc.(Eng.), A.C.G.I.†

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

Summary: The paper describes the operational tasks which will have to be carried out by a radar data handling system associated with an air traffic control centre—for example track initiation and prediction, data link interpretation, compilation of the air situation, aircraft control and safety, etc. Based on these considerations, the specification of a computer to perform these functions is discussed.

Consideration is given to the possible use of analogue and special-purpose computers but it is concluded that these are not suited to the task. Instead, a general-purpose digital computer would be required incorporating some special instruction facilities to assist the particular types of processing involved. This computer would use random access core storage for data and, possibly, program, although some form of permanent storage for the latter must also be provided. It is also shown that for good reliability and availability redundancy must be built into the system together with monitoring and maintenance aids.

1. Introduction

This paper will discuss the desirable characteristics of computers required to be used in a radar data processing system. It is intended to describe first some of the operational functions that the system will have to perform in an air traffic control centre. The centre, it will be assumed, is served by surveillance radars and is in communication with neighbouring areas. The aircraft within the area, it is assumed, will be the full range of civil aircraft with velocities up to just subsonic, together with military aircraft which may well be flying and manoeuvring at supersonic speeds; these latter aircraft, while not of direct interest to civil traffic controllers, cannot be ignored if only for reasons of safety. It is because of the wide range of velocities of modern aircraft, and the fact that with these high velocities aircraft are within the control area for only a short time during which vital decisions have to be made, that it is necessary to consider the introduction of automatic aids.

The later sections of this paper will describe the type of computer and its facilities that is considered to be required to satisfy the data processing requirement. Questions of reliability and maintenance aspects are also considered in some detail.

2. Radar Data Processing

2.1. *A Typical Radar Data Processing Environment*

Some of the tasks that a computer will be required to carry out in the environment of a radar data handling system are described in this section. It is assumed here that the computer is being employed to process data originating from an automatically detected radar source; such an automatic detection system is described in detail in a companion paper.‡ Briefly, as the radar beam sweeps across a target a number of returns will be received, the number being dependent upon the radar aerial beamwidth and the pulse repetition frequency. The successive returns from the same target during any specific radar rotation are processed by auto-detection equipment and the only information passed to the data handling computer is that equipment's assessment of the actual position of target source in range and bearing. Even so the quantity of information may be quite large. In the South of England, for instance, it is not uncommon to see over 100 individual tracks simultaneously displayed on a p.p.i. and the majority of these are presented to the data handling system on each rotation of the radar aerial (i.e. every 5–10 seconds, say). In addition there will be further detections, particularly from non-m.t.i. radars arising from clutter, e.g. clouds, land masses etc., which may well double the source detection rate. From this mass of detections presented to the data handling computer it will be necessary to attempt initiation and maintenance of the real tracks on the basis of correlation between detec-

† Admiralty Surface Weapons Establishment, Portsdown, Hampshire.

‡ J. V. Hubbard, "Digital automatic radar data extraction equipment". *The Radio and Electronic Engineer* (*J.Brit.I.R.E.*), 26, No. 5, pp. 397–404, November 1963.

tions from successive scans of the radar, to differentiate from the clutter sources and probably to display the real tracks, free from clutter returns, to the human operators.

2.2. *Track Initiation*

Considering specifically track initiation and track maintenance aspects, once a track has been initiated—i.e. at least two detections from different aerial rotations have been correlated together—velocity information is available; initially this velocity information will be of low accuracy, being based on only two detections each subject to system error. Prior to this, however, in the absence of any knowledge of course and speed, it is evident that a detection which might correlate with an earlier isolated detection could lie anywhere within an annulus drawn around the first detection, where the radii are governed by the time since the first detection. The outer radius is a measure of the distance that a maximum velocity aircraft would travel in this time whilst the inner is determined by the minimum speed of interest. Unfortunately, since allowance has to be made for radar errors and aircraft minimum velocities which may be less than 50 knots (ground speed) the minimum radius of the annulus may be considered to be zero. Although this simplifies the calculation, it does remove a possible means of distinguishing between clutter and aircraft. Thus any new detection found within the full circle centred on the original detection, and of radius controlled by the time which has elapsed since the earlier detection and the maximum speed of an aircraft in which we have an interest (in an air safety context this means all aircraft, military and civil alike), can be said to constitute a possible correlation on the basis of which a possible track may be initiated.

Consideration must be given to the period for which one should retain an uncorrelated detection in the data handling system in the hope that a further detection will occur with which it may correlate to provide a track initiation. Since at the commencement of a track, detections will be made when the probability of 'paint' is fairly low, i.e. 50% or less, it is essential to retain a detection for correlation for a period greater than that of one radar scan. If, however detections are retained for many scans, the size of the area in which a correlation detection may be present becomes very large and false correlation more probable. Furthermore the accumulation of uncorrelated detections will increase with time, posing a storage problem apart from the increase in computation time required to process the additional information. Once again a compromise must be sought and an adequate one would seem to be that uncorrelated detections should not be held for more than two or perhaps three rotations of the radar antenna.

2.3. *Track Maintenance*

Considering the case of an initiated track it is again necessary to consider all new detections that fall within a given distance of the existing track in order to identify the probable current position of the track. In this context, however, it is only necessary to look within a somewhat smaller area for the correlating detections since, using the previously calculated velocity, it is possible to predict the probable position of the track at the instant of the next detection and set the area limits from this position. The area to be inspected must, however, be one sufficiently large to allow for radar errors, inaccuracies in velocity components calculated at the time of the previous detection, and displacements which would result from the aircraft performing a maximum manoeuvre throughout the period of time since the last detection was made. If the environment is one involving co-operating civil aircraft only, provision for maximum manoeuvre should not be necessary but, in fact provision has to be made in order to track, for safety purposes, any aircraft, military and civil, not under the control of the centre.

2.4. *Correlation Ambiguity*

The question of ambiguities in the correlation of tracks with detections must be considered, and on the assumption that aircraft fly along approximately straight paths for a high percentage of their flight time, it is suggested that correlation of the detections closest to a track will, in most cases, result in correct association. However, because of the possibility of ambiguous associations it is not permissible to consider each new detection independently since tracks may be wrongly associated by so doing. Ideally all detections resulting from a scan should be considered together but this is probably inconvenient and not really necessary. Thus in general, all detections within a sector (say about 10 deg) should be considered at one time, all track/detection correlations within this sector being tabulated. Where there is no secondary radar agreement to eliminate the entries (see Section 2.5), this table should be inspected for the track/detection with the closest agreement and this correlation accepted. Acceptance of the correlation enables any other possible pairings using either this track or the detection to be discarded forthwith and, with the remainder, the process may be repeated. In this way the best solution for any given set of circumstances is achieved.

It may be thought that an impossibly high searching load is being imposed on the computer by this correlation process, but certain economies are possible. It is obviously not desirable to require the computer to search every existing track or uncorrelated detection each time a new detection is made by the radar and auto-detection equipment. If all tracks and uncorrelated detections were stored in order of one of their co-ordinates (i.e. range or bearing if processing is being carried out in polar co-ordinates, X or Y if processing is being carried out in cartesian co-ordinates) the search problem could be greatly eased. Such a table would have to be reshuffled for each radar scan in order to insert new uncorrelated detections and to correct the relative positions in the table of tracks whose co-ordinate has crossed another as a result of up-dating. The time required to reshuffle the table periodically is more than off-set by the adoption of a 'logarithmic' or 'binary' search of the table. The search is one in which the table is entered at its centre; the co-ordinate of the track found there is compared with the new detection co-ordinate, in order to determine whether any associated track is in the upper or lower half of the table. Depending upon this result the new entry is then compared with the centre of the appropriate half of the table. This process is repeated until the track is located whose co-ordinates are closest to the detection. The number of entries required to do this is equal to the power to which two must be raised to equal the number of entries in the table. Having located this particular entry it is necessary, of course, to move up and down from this point in order to identify all the entries that are within the area of error allowed for the predicted position as against the detected position.

2.5. *Secondary Radar*

It is evident that if all aircraft were equipped with secondary radar in which codes were used, which uniquely identified the aircraft, the general problems of track initiation and maintenance would be greatly simplified. No longer would it be any great embarrassment to have large boxes in which to search for correlating detections in order to initiate or maintain tracks; no longer would it be necessary to take measures to avoid incorrect correlation and to eliminate ambiguity. Comprehensive tracking programs would have to be retained however, to allow for the breakdown of the secondary radar in individual aircraft as well as to take account of aircraft not fitted with this equipment.

2.6. *Track Prediction*

Following the decision that a particular detection correlates with an existing track it is necessary to consider the position and velocity components that are to be used for future prediction of the track. So far as position is concerned it is evident that the new detection position must be accepted, although perhaps modified, to allow for the radar error of the detection source.

The velocity components to be used for further prediction must be adjusted on the evidence of the new detection. It is evident that some care should be taken to damp out velocity changes which arise from radar errors of the system, as such errors could give the impression that an aircraft is flying a continually varying instead of a straight course. Thus it is suggested that apparent small insignificant changes in component velocities should be damped out. It is important, of course, in considering future velocities to be used for track prediction, that they should reflect any control instructions passed to the aircraft.

2.7. *Data Links*

All the items considered so far in this Section have been concerned with picture compilation. To complete his picture the air traffic controller requires to know the air situation by detecting sources in neighbouring areas. Such information will facilitate easy transfer of control from one area to another and, since with their present-day velocities aircraft may remain within an area for a relatively short time, this information is almost essential to facilitate the long-term planning required to control all aircraft economically.

To obtain this information efficient data links capable of passing aircraft positional information, flight plan, secondary radar codes and aircraft identity are necessary. Passing the information with sufficient accuracy and detail by voice is not practical in the time available and the provision of digital data links becomes a necessity. These links should be automatically controlled, provisioned and interpreted by the data processing system.

Information received at a centre via data links must be correlated with that received from own detection sources. By so doing tracks can be maintained from one area to another, reports being maintained by the station currently detecting.

2.8. *Aircraft Control and Safety*

By means of the processes described, a complete picture of the air situation has been assembled and this may now be used by a central computer for the preparation of control instructions to aircraft in the zone. From this data the computer could decide the most economic flight paths for aircraft to use and the most economic order of landing and take-off. In arriving at its decisions the computer would, of course, take into account aircraft endurance. If data links between the control centre and aircraft were available, instructions and information could be passed to and from each automatically, but in the absence of such a facility it will be necessary to display the data for onward transmission by voice link and to provide means of manually injecting additional information into the computer (see Section 2.9).

It would probably be uneconomic and impracticable to program the computers to take care of all eventualities in the field of aircraft control. The main virtue of the application of computer techniques is to take away from the man the burden of repetitive and routine operation so that he can concentrate on the overall situation and the problems created by the exceptional event. An associated display system is required to facilitate this role of the man in an otherwise automatic environment. Equally, there must be a discipline placed upon the designers not to use equipment uneconomically in an attempt to automate a task for which the man is perhaps more suited. The resultant system should be one in which man and machine are operating in a partnership, each assisting the other where that assistance is most needed.

Aircraft safety is a particular aspect where computer assistance can be of great value, but one where automatic decisions may be particularly dangerous. To perform this task, it is necessary to predict future movements of each aircraft in the system and to examine the data continuously for possible conflict so that early warning for avoiding action may be given. To decide which aircraft constitutes such a risk is undoubtedly a task to which computer techniques may be applied. The avoidance action however may not be suitable for computer calculation since, provided the pilot is made aware of the situation and is supplied with the appropriate data (e.g. relative range, bearing and height of the other aircraft in the area), he is possibly in by far the best position to decide upon the immediate action to take since the computer recommendation under these conditions may be undesirably influenced by inaccuracies of the serving system. This must be true if both aircraft are not under direct control of the centre since the possibility of having a computer-controlled collision is all too real.

2.9. *Presentation to the Human Operators*

It has been indicated above that man will still retain an important role in an automatic radar data processing system. It is necessary, therefore, to ensure that he is not only provided with all the information he requires but that it is presented in a form that enables him to carry out his task efficiently. A display system which is suitable for this purpose is described elsewhere† and will not be described further here. All the information for this display system is generated, initially, by the computers of the data processing system, the display system itself converting this into marker or alpha-numeric display as required. The information which it is required to inject into the computer system by manual means may be carried out by a manual injection unit, a form of which is also described in the paper on display systems, and will not be discussed further here.

† D. R. Jarman, "The display of automatically processed radar information". *The Radio and Electronic Engineer* (*J.Brit.I.R.E.*). To be published.

3. Computer Characteristics

In this section it is intended to discuss the essential characteristics of a computer to perform the operational functions that need to be carried out by an automatic radar data processing installation.

3.1. *Analogue or Digital Computer*

The computer to be used in the system has to be capable of handling large quantities of input data from various sources, which will arrive at irregular intervals and may be in coded form. Their quality may vary with the result that the computer will be required to assess its value before further processing is carried out. Once the data are accepted, the computer will be required to carry out sorting and associating operations with that already held in the storage equipment and, as has been described earlier (see Section 2), it will be necessary to consider a large and variable quantity effectively in parallel.

Analogue computers are almost invariably special purpose and broadly speaking are only suitable for single-channel data processing not requiring great precision of computation in which the data inputs are derived from a single source, are consistently of good quality and can be manipulated to produce the required solutions without the need for storage of process history or associative data. Digital computers, on the other hand, can economically store large quantities of data and perform satisfactorily the types of operations indicated above using data derived from multiple input channels and specified in varying codes. Because of this, there is little doubt that the computer to be associated with the radar data processor should be digital.

3.2. *Special or General Purpose Computer*

The task of air traffic control, whilst being well known in principle, is one in which the detailed requirements are constantly changing with the development of air traffic. At the same time there are some processes in the maintenance of tracks by the association of new data which are specialized and unlikely to vary. Completely special purpose machines in which the program can only be modified by changes to the equipment are unlikely to satisfy the requirements of the data processor, and hence the computer should be general purpose with an instruction code tailored to meet the special requirements of air traffic control. Furthermore, the greater flexibility provided by a general purpose machine in this context is unlikely to result in the construction of a computer larger than the special purpose machine designed to carry out the same task.

3.3. *Special or Parallel Arithmetic*

The solution of air traffic control problems involves the computer in a considerable amount of routine arithmetic. This is particularly true of the association of incoming data with existing tracks, a book-keeping form of activity, and also of the routine up-dating of tracks. The need to carry out these activities in real time coupled with the variable rate of arrival of data means that a parallel arithmetic unit will almost certainly be required to cope with peak periods unless very fast computer logic and storage circuits are available.

3.4. *Word Length*

From the various tasks described, it is evident that there are certain to be a number of differing quantities associated with each stored track. These include range and bearing from the centre, (or X and Y if a cartesian co-ordinate system is to be used), rate of change of these co-ordinate quantities (representing the element of velocity), secondary radar codes, call sign of the aircraft and flight plan data (including for example expected time of arrival, endurance, etc.). None of these would appear to be quantities requiring a very large number of bits to provide an adequate accuracy. Taking range, for example, its maximum value with respect to any centre having an interest in the track is unlikely to exceed 500 miles; furthermore an incremental value of $\frac{1}{4}$ mile is the best that might be achieved from the detection sources. This then implies an 11-bit quantity plus a sign bit. However, the section on track correlation mentioned the need to predict the movement of tracks using calculated velocity components of the co-ordinate system. In order to avoid cumulative inaccuracies in this process it is necessary to hold range to a greater accuracy than that which it is possible to specify by a detected return from a track. These considerations indicate that a word length of some 18–20 bits, including sign, is required to satisfy this need.

3.5. *Storage Capacity for Data*

To assess the storage capacity of the data processing system it is first necessary to stipulate the maximum track capacity of the area in which the computer is likely to operate. In Section 2.1 it was noted that the simultaneous existence of 100 tracks surveyed by a single station in the South of England was fairly normal, and in Section 2.7 the desirability of accepting, via data links, track information detected at neighbouring centres was discussed. It is evident, therefore, that the system may well have about 300 aircraft operating within its area of interest. To store all the information required by the system on each of these tracks it would be necessary to consider the data required to maintain each track in greater detail than is possible in the space available in this paper, the significance to which each quantity is required, and the extent of which it is acceptable to pack unrelated quantities within a single word of the storage. However, it is fairly self evident that with the 20-bit words discussed in Section 3.4 above, an allowance of ten words would seem generous but should perhaps be used in the early stages of design in order to arrive at a safe figure for storage capacity.

Additional storage must be allocated to hold information on uncorrelated detections, correlated 'clutter' detections, working space, etc. Taking all these items into account it appears that 4000 words of storage would be inadequate while provision of 8000 words should provide an adequate margin of safety.

3.6. *Type of Storage*

Storage type should be considered under two headings, namely that required for data and that required for program. It is possible that different types would be desirable for each.

Data storage must be erasable and, for a parallel mode digital computer engaged on an air traffic control task, must have parallel random access. The requirement for a large amount of storage suggests, on a cost consideration, that a store of the familiar switched core should be chosen even if its cycle time does not match the speed of the arithmetic unit. In such a case the large capacity data store may be augmented by a number of fast computing registers, made from standard logic elements, which may be used to hold interim results of a computation and limit the use of the data store to providing initial data and storing final results.

Program storage may be erasable or fixed in the sense that it cannot be changed by either a computer fault or by electrical disturbances. If there is a need for frequent or rapid changes of program then erasable storage is the only practicable possibility but if the program is of reasonable length and does not require changes then fixed program storage may be preferred on the grounds of reliability. It is essential that a computer engaged in air traffic control, a task in which loss of control is dangerous, should have a program store which is both reliable and impervious to permanent damage from transient conditions. This indicates the need for a fixed program store. If requirements of program flexibility make erasable storage necessary then some form of back-up permanent storage should be provided. A magnetic drum which affords rapid reloading facilities, or magnetic tape, could satisfy this additional need.

3.7. *Instruction Code*

The tasks of the radar data processors described in this paper indicate that the computer will be required primarily to perform a sorting and correlating function on a large quantity of data generated in real time. To perform this and the other functions it could be easily demonstrated that the full range of conventional instructions must be provided, (for example, addition, subtraction, conditional jumps, shifts, etc.) but by more detailed consideration, it can be shown that certain functions not normally included in the instruction code of a general-purpose computer, are frequently in use and should be included if adequate capability can be built into the machine within practical computer speeds. To offset these additions simplified conventional instructions may, however, be acceptable. Some examples are given in the following paragraphs of these two factors.

(1) The correlating process requires repeated comparison of the co-ordinates of a new detection with those of existing tracks; the ensuing action depends upon whether these co-ordinates are within a certain value of each other. Since either co-ordinate can be the greater, and it is the magnitude of the difference that it is required, a modulus instruction would be of advantage. In this same context, the use of associative storage has application.

(2) The correlation process also requires similar computer action on the similar component of successively stored tracks. Since, as has been indicated, track information will occupy a block of storage, it seems necessary to provide facilities to modify and count in increments greater than 1 and which can be specified within the program.

(3) The use of marker bits has considerable application in data processing systems. Hence instruction facilities to address directly individual bits of the addressed word and to make their state the condition upon which future action should depend would be of advantage.

(4) It has been indicated that many of the quantities to be stored are of only a few bits significance. If storage is not to be used wastefully, therefore, it is desirable to pack more than a single quantity within one word of the store. In order to operate on these quantities speedily, special collating facilities should be introduced.

(5) The computer is intended to operate with automatic data links and it is probable that information on these links will be protected by some form of parity count; special instructions to generate parity will, therefore, be desirable.

(6) Although the computer is likely to operate in the parallel mode, the output to data links will probably be required in serial form. Consideration should therefore be given to provide circuits to facilitate conversion from serial to parallel mode and vice versa.

(7) The computer, being associated with a complex of external equipments (for example data links, auto-detection sources, teleprinters, etc.) must be provided with comprehensive interrupt facilities.

(8) In general, multiplication and division have only limited application. Thus special circuitry to provide fast operation of these functions is probably not worthwhile.

(9) Since the magnitude and scale of all quantities are known in advance there is no need to provide floating point arithmetic facilities.

4. System Design Considerations

Having selected the computing facilities required to carry out the air traffic control task, it remains to consider the system which can provide these facilities with an availability consistent with the task requirements. Clearly, the availability requirements of an air traffic control system are stringent for it must work in real time and failure may result in a dangerous situation.

At this point it is relevant to consider the basic form of the active equipment although, since the points to be raised under the headings of reliability and maintainability apply to all practical forms for a task as complex as air traffic control, there is no need to detail the various sorts of logic available. The advantage of cost, size, reliability and general suitability of semiconductors compared with electronic tubes makes it certain that the computer will be based on semiconductor techniques. In the semiconductor field, the constant introduction of new components and reduction in size of existing components coupled with new techniques for building them into circuits of steadily diminishing physical size are placing the portable computer among possible choices but at high cost. The air traffic control system, being ground based, should not be restricted by space limitations and equipment should be designed solely on an availability basis.

4.1. *Designing for Reliability*

System availability, which should always be specified in the initial design stage, depends on both reliability and maintainability. Reliability is usually measured in terms of the mean time between failures, but it should be noted that there is a difference between the practical and absolute reliability of equipment in the sense that if equipment can sustain failures and remain fully operational then such failures may be discounted in assessing practical reliability. Maintainability is a function of the mean time to locate and repair a fault.

Reliability, in the absolute sense, depends on the careful choice and use of components and on paying meticulous attention to system integration at all levels from the insertion of a component into its circuit to the interconnection of the computer with its major items of ancillary equipment. The achievement of consistency of reliability throughout the system should always be a major design aim.

Designing for reliability at circuit level includes the use of circuits which continue to function correctly in the face of substantial power supply and component value variation, circuits which fail safe, and circuits incorporating redundant and underrun components. All circuits should be tested under the worst expected conditions and possibly with known limiting value components included. The use of redundant and underrun components is, in effect, a means of buying reliability by using more and better components than are strictly necessary to do the job. Since the failure of a redundant component does not impair the operation of equipment it is an example of keeping 100% practical reliability in the face of component failures.

At system level, designing for reliability is concerned with the electrical matching of circuits to one another, the method of interconnecting them mechanically and the provision of standby units which either duplicate working units or can replace them functionally with perhaps a reduction of operational efficiency.

Starting with the computer there is undoubtedly a case in a task like air traffic control for dividing the computing load between several computers which can together provide the requisite computing power. This arrangement reduces the possibility of system failure if a means of keeping the system operational in the event of one computer becoming unserviceable is provided. One such means is to provide a complete standby computer although this is expensive and could only be justified if a large number of very small computers were used or alternative work outside the system could be found for the standby. A more practicable proposition is the use of, say, three computers any of which, by interchange of program storage, may perform any of three computing tasks which together comprise the whole air traffic control task. In this event the standby arrangements, which could well be wholly or partially manual, are required to cope only with the least important computing task.

Having settled the computer complement, attention must next be paid to the provision of standby equipment for units which are smaller than computers though still major items. Standby equipment for such units are virtually indispensible where several computers rely on, say, a relatively small input control unit. Such a unit is an obvious candidate for duplication and it is usual to feed both the working and standby units with the operational inputs and compare their outputs. An independent check is required on correct operation, of course, but this procedure ensures that the standby unit is functioning properly and that it is in the correct state for insertion into the system.

The third level of standby units entails providing redundant circuits, either directly or in the form of voting logic, in major equipment units. The practice in the direct application consists, typically, of placing circuits in parallel where one circuit, in failing safe, cannot impair the correct operation of the other. These arrangements are generally made in important sections of the system or in sections where checking for correct operation or fault diagnosis is difficult. Since no fault will be apparent in the event of one circuit failing it is necessary to carry out checks on redundant circuits on a routine basis. In voting logic important decisions are made by the agreement of, say, two out of three units or circuits operating in parallel. The principle is an extension of direct redundancy, and the same remarks apply, but it may be applied successfully, whilst being more expensive in equipment, in a wider variety of situations.

To sum up, designing for reliability requires the provision of an environment defined by published data on components and the provision of adequate standby facilities. It is essential therefore for the design to include the provision of this environment and to ensure that the best possible use is made of the standby facilities.

4.2. *Environmental Testing*

The problem of electrical environment is an integral part of circuit design with which every designer is familiar, but the problem of physical environment is of equal importance. Physical environmental stability includes the maintenance of suitable temperature and humidity, and freedom from corrosive agents, shock and vibration. Temperature control provides stability in circuits and in the case of semiconductors, for example, allows design for reliability by underrunning to proceed on a realistic basis. The designer must, of course, take care not to negate the benefits of controlled ambient temperature by allowing circuits to create local hot spots in critical places.

An aid to achieving operational reliability is the provision of the means of increasing stress on the circuit elements in order to check that the system has a margin of safety. These checks are usually electrical, although heat and vibration checks are also a possibility, and are commonly called marginal checks. They are applied to equipment on a routine basis during maintenance periods and are intended to stress components beyond their normal working limits (though not, of course, in a way likely to damage them) and to make those which are on the point of failure due to characteristic degradation fail at a time when no harm can be done to operational activities. Marginal checks must be applied with care since there is no point in creating faults which cannot be traced and repaired, and it follows that an efficient fault finding capability is an essential supporting facility to marginal checking. When faulty components have been replaced the removal of margins gives a reasonable assurance that no more component failures due to degraded characteristics are imminent.

Finally to cater for those sections of equipment to which marginal checking cannot be applied there is the use of a routine replacement policy. This entails the replacement of components or units at fixed intervals determined by component life test results, electrical and mechanical environment, and experience. Routine replacement must be applied with discretion since there is evidence supporting the view that reliability is improved by leaving well alone, but it is useful in the case of electromechanical devices, the life expectancy of which can be assessed with reasonable accuracy, and where the incidence of failures of specific components or components in specific locations suggests that replacing all similar components would be beneficial.

4.3. *Standby Replacement Units*

Turning now to the role of standby units it is clear that such units must always be checked, ready for service, easy to connect into the system and preferably with information content operationally up-dated. Further, it is essential that prompt warning of system failure is given to indicate that a standby unit is required. In other words the system must have self-checking and fault-indicating facilities. These facilities consist typically of computers checking themselves by performing calculations on test data, giving a fault warning if the known correct result is not obtained and checking other units by testing their outputs at times in the data processing cycle when either these outputs are already known or can be forced to a known quantity by injecting test data into the units.

The change over to a standby unit can often be carried out automatically by equipment which responds to a failure signal. This principle may be extended to the automatic isolation of units which play a secondary role and can be discarded if faulty. Little has been written about equipment ancillary to the computers except to state that it must conform to the overall system availability requirement, but it should be noted that these units can also play a part in the system checking and correcting activities. The services in the automatic fault correcting category are provided at component and elementary circuit level by redundant components and circuits as described above, but may be extended, in addition to the automatic change over to standby units, to tasks such as the automatic recharging of erasable program stores, the content of which has been found faulty by, say, a parity check, from a backing up store. Where data inputs to a system are found to be faulty then either automatic isolation or, less frequently, automatic correction may be provided. Automatic correction is usually used where data generated at a distance from the processing site are prone to transmission error. The data are subjected to comprehensive parity encoding before transmission which allows the receiving equipment to check if the data are correct and to correct them if the number of errors is small. Data which cannot be corrected are rejected by the receiving equipment. It is also desirable to provide means of automatic isolation of faulty data sources to prevent the accumulation of false data in the system. A system can often remain effective when deprived of one of several sources of primary data or of second order refinements to primary data.

4.4. *Maintainability Considerations*

The basic requirements of reliability have now been covered and, having tried to ensure that the system will not fail, it is time to consider what to do when it does fail and how to arrange that a faulty unit replaced by a standby unit is quickly repaired and ready for further service. The key to these problems is the provision of good maintainability, which means the rapid location and repair of faults. A general purpose digital computer is a very powerful tool for the diagnosis of faults although the preparing and checking of procedures which are both comprehensive and unambiguous is long and tedious. They are, however, a virtual necessity for computers engaged in a task like air traffic control. It is common for digital computers to diagnose faults within the major part of themselves by program provided that the part of the computer required to run the program is fault free; faults on this part may be diagnosed by an external unit specially designed for the task and connected in as required. It is also possible for a computer to diagnose faults in any other unit with which is has two-way communication by sending test data to it and analysing its outputs. Further, a correct computer can monitor a faulty one if they can run the same program and communicate with one another to compare progress. This is not usually a practicable proposition however since a system can seldom dispense with two computers to engage in fault finding activities.

In considering ways of achieving rapid and effective fault finding there can be no doubt that one of the key factors is equipment simplicity. In the logic of a digital computer one of the main causes of complexity is making circuits common to several different functions, a practice which often results in fault symptoms which are difficult to analyse. In this case simplicity results from using more circuits to separate the functions, a course which will increase the hardware required and reduce reliability simply because there are more components, but the net gain in equipment availability is usually worth having.

Another important factor in making fault finding easy is the choice of content of replaceable units. The subject of replaceable unit size will be considered later, but whatever the size of the unit its content is primarily determined by the need to make it suitable for *in situ* checking. This requirement can best be met by making the unit perform a small number of easily checked functions. Unit content is also determined to some extent by the need to keep the number of different unit types to a minimum but this requirement is less important than ease of fault finding. It should be noted that there is often very little merit in using large numbers of units each containing an identical simple circuit since, although the circuit function may be easily checked, it is difficult to arrange diagnostic procedures which are effective at single logic element level.

Having found a faulty unit the next problem is how best to repair it and this involves the choice of unit size, the method of connecting units into the main equipment and accessibility. Unit size is unimportant from a replacement point of view unless the unit is unmanageably large or accessibility is poor. There is no doubt that for much of the equipment, especially the computers themselves, the policy of repair by unit replacement is essential, but it does not follow that treatment needs to be uniform over the whole system. If, for example, a system, for which the overall policy is to employ small sub-units, includes major units, for which standby equipments are fitted, it may be better to use whole unit construction for these, thereby gaining reliability by dispensing with plugs and sockets, and to repair a faulty unit *in situ*.

If replaceable units are to be used the method of connecting them into the main equipment is determined by the ease and speed with which units must be replaced. Here, the price of ease and speed of replacement, which usually means the use of plugs and sockets and possibly non-rigid retention in cabinets, is reliability because plugs and sockets are by no means the most reliable components.

There is little need to dwell on the topic of accessibility except to stress that it is of great importance and that the designer who is lucky enough to be entrusted with the mechanical design of a ground based equipment, as is the case with an air traffic control system, has little excuse for not providing optimum conditions for maintaining the equipment.

The final topic is that of repairing replaceable units. This should always be carried out in test equipment external to the main equipment and the most important requirement is that the test set should simulate the true environment of the unit closely enough to ensure that if a unit is passed by the test set as satisfactory it can be returned to the equipment with certainty that it will function correctly. Test equipment is expensive and complex if the number of unit types is large and they require a large number of complex inputs and produce complex outputs, but nevertheless the units should always be chosen on the basis of suitability for the main equipment rather than of suitability for repair.

5. Conclusions

This paper has endeavoured to define the specification of a computer which would be required to satisfy the needs of a radar data processing system associated with an Air Traffic Control Centre. In order to satisfy this specification a brief description of some of the major tasks that the computer may be required to carry out have been described. From these considerations it has been concluded that the system will best be served by a digital general purpose computer incorporating some special instruction facilities to assist the particular type of processing involved. To achieve the necessary speeds with currently available components, parallel arithmetic would be required and use made of random access core storage for data. Program storage of a permanent form would be required in this system although it may not be considered necessary to use such storage directly by the computer control but only as a form of back-up. For reliability, the use of redundancy in some form—possibly indirectly as low capacity operational ability—is necessary, while to achieve high availability monitoring and maintenance aids are an essential need.

6. Acknowledgments

This paper is published by permission of the Board of Admiralty. The authors wish to acknowledge the contributions of their colleagues at A.S.W.E. to this work.

Manuscript received by the Institution on 1st April 1963. (*Paper No. 844/RNA23*).

(Discussion on this paper appears on page 416)

DISCUSSION

Under the Chairmanship of Mr. R. N. Lord

Mr. P. N. G. Knowles: An automatic tracking system often requires reasonably accurate velocity data both as an output and as an aid to the tracking process. Mr. Ballard suggested a range quantum size of $\frac{1}{4}$ to $\frac{1}{2}$ mile. This could give sufficiently large random velocity and acceleration errors to make tracking difficult. I would suggest that $\frac{1}{16}$ mile quantum size is the minimum size acceptable for the tracking of current typical aircraft

The reply was that finer granularity was used in the machine described for velocity extraction.

The Authors (*in reply*)*:* The significance of $\frac{1}{4}$ to $\frac{1}{2}$ mile range quantities was referred to in the context of the presentation on instantaneous position. The paper agrees with the questioner that, to permit accurate up-dating of tracks, range information must be stored to a greater accuracy than this. It was primarily for this reason that it was concluded that a data word of some twenty bits in length would be required in the computer.

Mr. F. Barker: Speed of multiplication: Since scan to scan correlation of track data is an important function of the computer, and since data on individual plots will normally be yielded in polar co-ordinates, it would appear that fast multiplication would be useful for the rapid conversion of this data to rectangular co-ordinates, in which terms correlation is most easily effected.

The significance of manual inputs: Although the system with which the paper is concerned is largely automatic, it is probably true to say that, at the present time, any system for Air Traffic Control would be required to accept a large number of operator inputs and in real time. It is therefore important in considering the type of computer, to study how it can best be orientated towards these manual inputs which have their own problems, for example their relatively wide variety and random order of arising—quite rapidly in a large system.

The Authors (*in reply*)*:* It is not necessarily true that correlation can best be carried out in cartesian co-ordinates but accepting this statement it does not necessarily follow that fast multiplication is an essential facility of the computer. It must be remembered that the number of significant bits of the quantities to be multiplied in co-ordinate conversion is not large and, as a result, there is no need for full word fast multiplication.

With reference to the significance of manual inputs, it is agreed that the system must be designed bearing in mind that not all information can be derived from automatic sources and that manual input facilities must be provided and integrated with the system design.

U.D.C. 621.3.011.222: 681.33

A Series Type D.C. Negative Resistance for Analogue Computers

By

P. V. INDIRESAN, Ph.D., (*Associate Member*)†

AND

Mrs. SASHIBALA SATAINDRA, B.E.†

Summary: Negative resistances are only conditionally stable, the shunt type when connected to a positive resistance of a smaller value and the series type when the load resistance is larger. Hence, in general, in an electrical resistance analogue, both series and shunt types of negative resistance will be required. Methods of obtaining these two types of negative resistance for operation at both d.c. and a.c. are described. The principle, design and operating limits of the series type of circuit are discussed in detail.

1. Introduction

A number of problems involving the computation of various types of fields can be conveniently solved by the technique of electrical analogy. Except for the simplest cases, the electrical analogue will need both positive and negative circuit elements. Inductors and capacitors have been used for this purpose in a.c. analogues,[1, 2] but their use is limited because (*a*), the parasitic coupling between inductors is difficult to overcome, and (*b*) it is expensive to match the components to the requisite degree of accuracy. Alternatively, cascaded networks[3] can be used but they too are not convenient because a large number of essentially ideal transformers will be needed and also because unwanted phase shifts create complications. For these reasons a.c. analogues have found only a limited use in practice.

However, these difficulties of parasitic coupling and phase errors can be entirely eliminated by using a d.c. analogue with positive and negative resistances in place of the a.c. analogue employing reactances. The conventional negative resistances are not suitable for this purpose as they do not operate with direct currents. Theoretically, negative resistance elements like tunnel diodes and dynatrons can be used in conjunction with appropriate biasing potentials, but this is impracticable as the bias adjustment is critically dependent on external load connections. The same difficulty is encountered with the negative resistance circuit devised by Swenson and Higgins.[4] In these cases, the adjustment of each negative resistance affects the balance of all the others in the network so that they all have to be corrected iteratively. This type of correction fails to converge for analysis above the first anti-resonant frequency of the system[5] and therefore the scope of the analogue is severely restricted. For these reasons, the negative resistances used in d.c. analogues should not only be capable of operating at d.c. but should also be independent of external load connections. In addition, as they will be needed in large numbers, it will be advantageous if they are economical in cost and compact in size.

A shunt type of transistor negative resistance having all these characteristics has already been described[6] and has been very successfully used by Redshaw and Rushton[7, 8] for the solution of problems in elasticity. However, a shunt type of negative resistance is stable only when it is shunted by a smaller positive resistance. Generally this condition cannot always be met, hence it is necessary to have a series type of negative resistance in addition to the shunt type.

This paper describes a simple, economical transistor circuit which provides a series type of negative resistance whose magnitude is unaffected by the external connections. This series-type negative resistance is complementary to the shunt type already described and the two together should make almost any type of resistance analogue electrically stable and practicably feasible.

2. The Shunt Type of Negative Resistance

Figure 1 shows the practical form of shunt type of negative resistance circuit. It is essentially a bistable

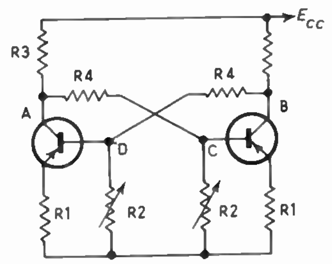


Fig. 1. The shunt type of negative resistance circuit.

† Electrical Engineering Department, University of Roorkee, India.

multivibrator and the negative resistance appears between the terminals A and B. The magnitude of this negative resistance R_n is given approximately by

$$\frac{2}{R_n} = \frac{1}{R_3} - \frac{1}{R_1} \cdot \frac{\alpha R_2 - R_1}{R_2 + R_4} \qquad \ldots(1)$$

When no external voltage is applied the potentials at the terminals A and B should be balanced and this will happen only when the two transistors and the resistors are exactly matched. However, very precise matching can be avoided by making the resistors R2 variable. This also helps to adjust the magnitude of the negative resistance. The voltage range of operation of the circuit is increased and the unbalance voltage is reduced when R_1 and R_4 are kept large or when R_2 is made small. Varying the load across the negative resistance does not at all affect either its magnitude or its voltage range of operation. As the load resistance approaches the value of the negative resistance the adjustment of balance becomes more and more critical. A detailed description, analysis and the principles of design of this circuit as well as a discussion of the stability of negative resistances is given in detail in reference 6.

3. The Series Type of Negative Resistance Circuit

The circuit of Fig. 1 itself gives a series type of negative resistance between the points A and C or between B and D. It is, however, quite impractical as it is very difficult to balance the voltages at these points. This difficulty is avoided in the circuit shown in Fig. 2. The action of this circuit may be explained as follows.

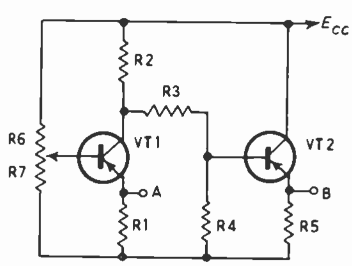


Fig. 2. The series type of negative resistance circuit.

When a current is passed from an external source from A to B, the voltage at A becomes positive. The transistor VT1 acts as a grounded base amplifier and consequently its collector becomes even more positive. This amplified positive voltage is applied to the base of the emitter follower VT2, through the potential divider formed by R3 and R4. If the overall amplification is greater than unity, the voltage at B becomes more positive than that at A. In other words, when the current flows from A to B, B becomes more positive than A and hence the effective resistance between A and B is negative. It is evident by inspection that this circuit is stable when A and B are open circuited and unstable when the two are short circuited. Therefore, this is a case of an open circuit stable, current actuated, series type of negative resistance.

4. Magnitude of the Negative Resistance

If $-R_n$ is the magnitude of the negative resistance, the circuit will be stable when the load across A and B is greater than R_n and unstable when it is less. When the load equals R_n the circuit is critically stable and the loop gain round the feedback circuit just equals unity. This fact can be utilized to determine the magnitude of the negative resistance.

If A_1 = gain of the common base amplifier VT1,

A_2 = fraction of the collector voltage of VT1 applied to the base of VT2,

A_3 = gain of the emitter follower VT2,

and R_{11} = input resistance of VT1,

then, $$A_1 \cdot A_2 \cdot A_3 \cdot R_{11}/(R_{11}+R_n) = +1 \qquad \ldots(2)$$

But $A_1 = R_L/R_{11}$

$A_2 = R_4/(R_3+R_4)$

$A_3 = 1$

and $$R_{11} = r_{11} \cdot R_1/(r_{11}+R_1) \qquad \ldots(3)$$

where R_L = effective load resistance for VT1

$= R_2(R_3+R'_4)/(R_2+R_3+R'_4)$

R_{12} = input resistance of VT2 $= R_5/(1-\alpha)$

$R'_4 = R_4 \cdot R_{12}/(R_4+R_{12})$

$$r_{11} = r_e + \left(r_b + \frac{R_6 R_7}{R_6+R_7}\right) \cdot (1-\alpha) \qquad \ldots(4)$$

Substituting eqns. (3) and (4) in eqn. (2),

$$R_n = R_{11} - \frac{\alpha R_2 R'_4}{R_2+R_3+R'_4} \qquad \ldots(5)$$

If R_1 and R_5 are large, the negative resistance value becomes approximately

$$R_n = r_{11} - \frac{\alpha R_2 R_4}{R_2+R_3+R_4} \qquad \ldots(6)$$

5. Experimental Results

In Table 1, the values of negative resistances obtained experimentally for various combinations of the circuit resistances are compared with the theoretical values calculated from eqn. (6). It may be noted that this approximate equation forms an adequate starting point for the design of the circuit. From the approximate values of the circuit parameters so obtained, the adjustment for the exact value of the negative resistance can be made by varying R_2 or R_4. The voltages at A and B are balanced by means of the potentiometer R5, R6. An idea of the relative effects of these parameters can be obtained from Figs. 3 to 7 which show typical variations of R_n and of E_{max} the maximum

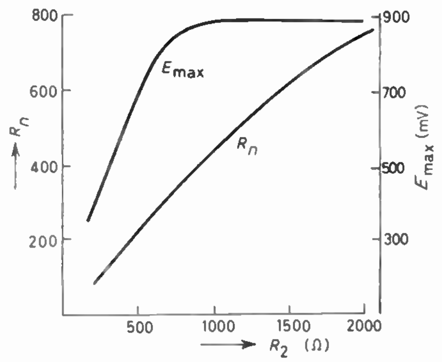


Fig. 3. Effect of varying R_2. $R_1 = R_5 = 1000$ ohms; $R_3 = 2000$ ohms and $R_4 = 3000$ ohms.

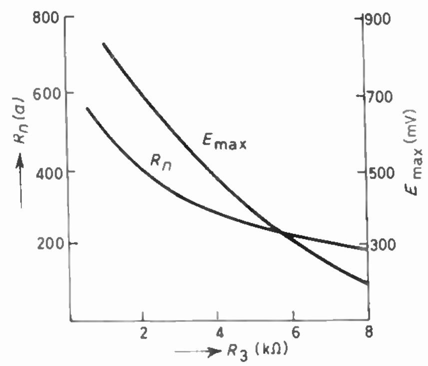


Fig. 4. Effect of varying R_3. $R_1 = R_2 = R_5 = 1000$ ohms and $R_4 = 3000$ ohms.

Table 1

Verification of the formula for the magnitude of negative resistance

($R_1 = R_5 = 1000$ ohms; all values in ohms)

R_2	R_3	R_4	R_n calculated	R_n experimental
200	2000	3000	82	80
500	2000	3000	235	213
800	2000	3000	371	359
1000	2000	3000	455	416
2000	2000	3000	797	742
1000	1000	3000	554	520
1000	3000	3000	384	337
1000	5000	3000	291	263
1000	8000	3000	210	182
1000	2000	1000	213	193
1000	2000	2000	349	337
1000	2000	5000	572	540

range of operation, when R_2, R_3, R_4, and the supply voltage are varied.

As expected R_n and E_{max} were found to be totally independent of the load resistance subject to the condition that it was not small enough to make the circuit unstable. Figure 6 also indicates the manner in which the balance adjustment varies as the supply voltage is varied.

The frequency range of operation of the circuit is limited essentially by the characteristics of the grounded base amplifier VT1 and the limit is set by a stray inductive reactance that invariably occurs in series with a series type of negative resistance. For the transistors used (OC72) this reactance was negligible at audio frequencies equalling the value of the negative resistance at about 85 kc/s. By using transistors with a higher cut-off frequency, a correspondingly higher frequency range can be obtained.

6. Comparison between the Shunt and Series Types of Negative Resistance Circuits

Normally, the use of the series and shunt types is mutually exclusive; when the shunt type is stable the series type is unstable and vice versa. However, it may at times be possible to choose the scale factors in such a manner that either type can be used. In such a

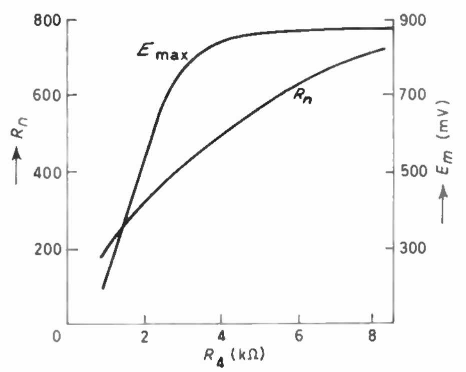


Fig. 5. Effect of varying R_4. $R_1 = R_2 = R_5 = 1000$ ohms; $R_3 = 2000$ ohms.

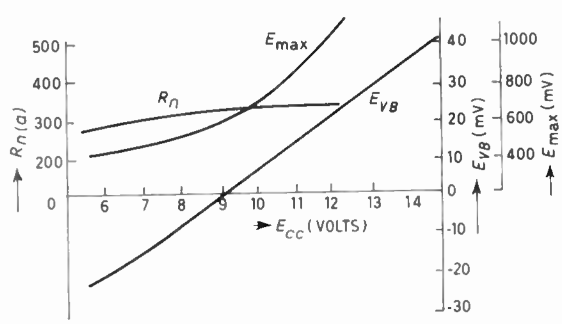


Fig. 6. Effect of varying the collector supply voltage.

case, it is of interest to know which type should be preferred. The *series type* has the advantage that

(*a*) it requires one resistor less and thus is more economical,

(*b*) as it is used with larger resistors the power consumed from the supply to the analogue is less and

(*c*) it does not need matched pairs of components.

The *shunt type* has the advantages that

(*a*) being a symmetrical circuit, the balance is maintained more easily,

(*b*) for a given voltage range of operation it requires a smaller collector supply voltage and hence consumes less power from the collector supply and

(*c*) the magnitude of the negative resistance is totally independent of the collector supply voltage. The advantages of the shunt type outweigh those of the series type, hence where possible the shunt type should be preferred. It must be emphasized that in either case the circuit balance adjustment becomes

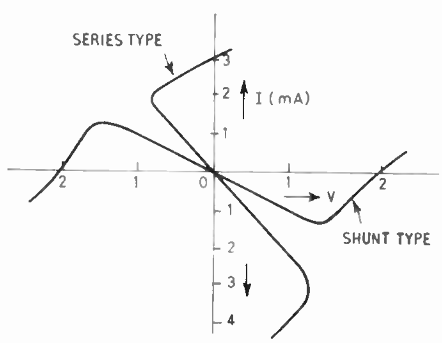


Fig. 7. Comparison of the characteristics of the series and the shunt type of negative resistance

more and more critical as the load resistance approaches the value of the negative resistance. For this reason, as far as possible it must be so arranged that every negative resistance used differs by an appreciable ratio from the load across it. In Fig. 7, the typical voltage current characteristics for the two types of negative resistance are shown. It may be noted that while the characteristic of the shunt type is symmetrical, that of the series type is not. The series type has an S-type of characteristic and the shunt type an N-type of characteristic.

7. Conclusion

One shunt and one series type of negative resistance circuit that can operate at d.c. as well as at a.c. has been described. Both have values independent of external load connections, but the former is stable when shunted by a smaller resistance and the latter requires a load of larger value for stability. Thus, by proper choice of the type of negative resistance, electrical stability can be obtained whatever the relative values of the load and the negative resistances. The circuits are particularly convenient for use in the large numbers required in electrical analogues as they are economical in cost, compact in size and simple to adjust. In fact, the shunt type has already been used successfully for the solution of problems in elasticity. Normally the use of each type is exclusive of the other because when one is stable the other is not. However, when it is possible to design a system where either type may be used, the shunt type appears to be preferable as its adjustment is relatively stable, its power consumption less and its magnitude is independent of variations in the supply voltage.

8. Acknowledgments

The authors are grateful for the encouragement given by Professor C. S. Ghosh and Mrs. Sataindra thanks the Council of Scientific and Industrial Research, New Delhi, for the research fellowship awarded.

9. References

1. G. Kron, "Equivalent circuits for the elastic field," *J. Appl. Mech.*, **2**, No. 3, p. 149, September 1944.
2. K. Spangenberg, G. O. Walters and F. Schott, "Electrical network analysers for the solution of electromagnetic field problems, (Parts 1 and 2)", *Proc. Inst. Radio Engrs*, **37**, p. 724 and p. 866, 1949.
3. W. T. Russel and R. H. MacNeal, "An improved electrical analogy for the analysis of beams in bending", *J. Appl. Mech.*, **20**, p. 349, 1953.
4. G. W. Swenson and J. J. Higgins, "A direct current network analyser for solving wave equation boundary value problems", *J. Appl. Phys.*, **23**, p. 126, 1952.
5. G. Liebman, "Solution of partial differential equations by network analogue", *J. Appl. Phys.*, **1**, p. 92, 1950.
6. P. V. Indiresan, "A negative resistance for d.c. computers", *J.Brit.I.R.E.*, 19, p. 401, 1959.
7. S. C. Redshaw and K. R. Rushton, "Various analogues, incorporating negative resistances, for the solution of problems in elasticity", *J. Appl. Phys.*, **12**, p. 390, 1961.
8. S. C. Redshaw, "The use of a negative resistance for the solution of structural problems by electrical analogy", Third International Computation Meeting, Opatija, September 1961.

Manuscript first received by the Institution on 13th March 1963 and in final form on 16th May 1963. (*Contribution No. 71/C61*).

U.D.C. 537.224/6 : 621.3.028

Transient Analysis of Thin-Film Structures

By

F. A. LINDHOLM, Ph.D.†

AND

W. W. HAPP, Ph.D.‡

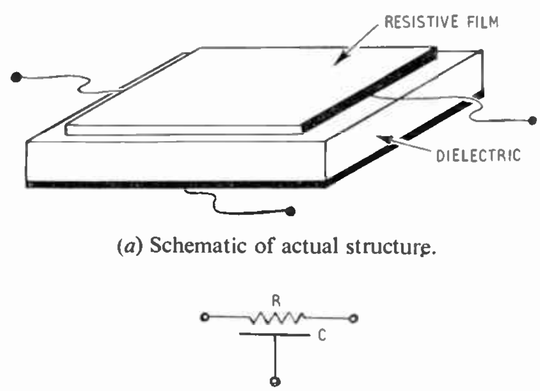
Summary: To provide the designer of digital circuits with the potential of a recently developed class of micro-system structures, consisting of thin films of resistance plated on a dielectric substrate, the transient behaviour of these components was investigated analytically and experimentally.

A survey was made of alternative approaches for computing the transient response. For arbitrary termination and drive, two methods proved effective: one primarily of mathematical significance, the other of more practical use to the circuit designer.

Following the second approach, a lumped model based on physical phenomena was developed and demonstrated to be optimum for circuit design. Experimental agreement with wave shape and response time was excellent.

1. Introduction

Small-scale circuits called for in micro-system electronics rule out the use of lumped resistors and capacitors. Phenomena occur on a differential basis: for example, current flow through a resistor becomes diffusion flow at a point; charge storage in an incremental volume replaces the conventional concept of charge on a capacitor plate; a network with a finite number of elements becomes a system with an infinite number of elements.



(*a*) Schematic of actual structure.

(*b*) Network symbol.

Fig. 1. Microsystem structure.

Many combinations of resistive films and dielectric films are used in practice. A representative micro-system structure, (shown in Fig. 1) consists of a wafer of dielectric material with a thin resistive film deposited on the top side and a conductive film plated on the underside. Electrical contacts are made to the ends of the resistive film and to the conductive film, making it a three-terminal device. This two-port network is a distributed parameter network[1, 2] conveniently denoted by the symbol shown in Fig. 1(*b*). Networks of this type are potentially important to the field of computer electronics not only because of volume reduction, but also because of their inherent increased reliability arising from fewer soldered connections. To assess the potential value of these networks as components in relaxation oscillators, pulse and logic circuits and others, their transient response must be examined. Although the frequency response of these networks is well understood,[3,4] little information on the transient behaviour is available.

A transient analysis useful for practical circuit design is the objective of this paper.

1.1. *Requirements for Circuit Design*

An optimum description of the transient response provides the circuit designer with

(*a*) an accurate response time,

(*b*) an accurate wave shape,

(*c*) a solution for arbitrary termination (source and load),

(*d*) a solution for arbitrary drive (voltage and current),

(*e*) reasonable accuracy with a minimum of mathematical calculation,

(*f*) reasonable accuracy with a minimum of graphical construction,

(*g*) a means to conventional circuit design,

(*h*) a means to construction of charts for routine design,

(*i*) a method displaying the essential physical processes,

(*j*) a method of acquiring knowledge for the evaluation of circuit behaviour,

† Assistant Professor, Department of Electrical Engineering, The University of Arizona, Tucson, Arizona.
‡ Research Professor, Department of Electrical Engineering, Arizona State University, Tempe, Arizona.

(k) a set of approximations of increasing accuracy,

(l) an estimate of error associated with each successive approximation.

The last two criteria deserve amplification. Circuit design frequently involves two steps: initially, an order-of-magnitude approach to assess feasibility of design, and subsequently, a more quantitative approach to make certain that the circuit meets specifications. Correspondingly, many design problems require models of two levels of accuracy: initially, a simple model for order-of-magnitude design, and subsequently a more adequate model for numerical evaluation. It is fortuitous that the simple model often gives results of adequate accuracy, thus eliminating the need for subsequent refinement.

1.2. *Analogous Systems*

It is advantageous to look for analogies to networks characterized by distributed resistance and capacitance. Since this type of network is governed by the diffusion equation, many systems[5] governed by this equation are analogues. In particular, the flow of charge in the distributed network is analogous to the flow of heat energy in a one-dimensional rod.

Two advantages result from exploiting this analogy: First, a conceptual clarification of the underlying physical phenomena is obtained. From a more practical point of view, the methods for solution of this classical problem,[6] including the powerful Laplace transform approach, are directly applicable to the distributed parameter network under consideration.

1.3. *Proposed Approach*

As a first step, the heat analogy is exploited by considering the method for solution of the heat problem which is most compatible with the requirements of circuit design. The Laplace transform yields an exact solution for arbitrary termination and drive, but only after considerable computational effort. Thus a search for a different approach is required.

Linvill[7] noted that concentration on the differential behaviour of a device (or, what is equivalent, analysis of the distributed model of a device) leads not only to prohibitive computational labour, but also to results that are incompatible with the demands of flexible circuit design. Motivated by Linvill's pioneering work on lumped models for diodes and transistors, a lumped model of the distributed-parameter network is constructed.

The versatility and power of the lumped model will be demonstrated by developing a design chart for use in repetitive design work and by predicting the transient response of micro-systems in typical applications. The transient response computed by this technique is then verified experimentally to assess the accuracy to be expected.

2. An Exact Solution by Laplace Transform

2.1. *The Method*

A systematic and effective method for obtaining the transient response of a distributed-parameter network terminated in an arbitrary source and in load impedances is due to Gray[8] who derived sets of calculated

Table 1

Physical laws of differential system

Physical Law	Thermal System	Electrical System	Equation
Diffusion flow	Flow of heat energy due to diffusion is proportional to the temperature gradient at point x and time t.	Flow of electric charge due to diffusion is proportional to voltage gradient at point x and time t.	$I_{\text{diff}}(x, t) = -A\dfrac{\partial V(x, t)}{\partial x}$
Storage	Heat energy stored in differential volume about point x at time t.	Electric charge stored in differential volume about point x at time t.	$Q(x, t)\mathrm{d}x = C\mathrm{d}x\, V(x, t)$
Storage flow	Time rate increase of heat energy stored in differential volume about point x at time t. Flow of heat energy to storage.	Time rate increase of electric charge stored in differential volume about point x at time t. Flow of current into storage.	$\dfrac{\partial Q(x, t)}{\partial t}\mathrm{d}x = C\mathrm{d}x\,\dfrac{\partial V(x, t)}{\partial t}$
Continuity	Net flow of heat energy into differential volume about point x at time t equals the time rate increase of heat energy stored in the same differential volume.	Net flow of electric charge into differential volume about point x at time t equals the time rate increase of charge stored in the same differential volume.	$\dfrac{\partial I_{\text{diff}}(x, t)}{\partial x} = -C\dfrac{\partial V(x, t)}{\partial t}$

curves for the transient response of a heat pump. Although this problem is not an exact analogue to the one under consideration, the similarity of the mathematical formulation involved warrants the use of Gray's method as a guide. In fact, this procedure was used by Starner[9] in a preliminary investigation of the transient response of thin-film structures. The analogy is shown in Table 1. It may be noted in Table 1 that the terms 'convection' or 'drift' are usually associated with the flow of electric charge arising from a voltage gradient. However, because of the analogy being drawn between temperature and voltage, it appears that 'diffusion' is the most appropriate term to use in the present investigation.

2.2. *Procedure*

For the geometry in Fig. 1 a transient solution in terms of voltage and current is obtained by this method in the following sequence.

(1) Solve the diffusion equation by assuming exponential variation with time for current (I) and voltage (V). Express these solutions in terms of the terminal variables following essentially the procedure proposed by Holm.[1]

(2) For the distributed-parameter circuit (Fig. 1) obtain the matrix equation:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{R}{\theta}\begin{bmatrix} 1/\tanh\theta & 1/\sinh\theta \\ 1/\sinh\theta & 1/\tanh\theta \end{bmatrix}\begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$$

where $\theta = (j\omega RC)^{\frac{1}{2}}$; it is desirable to extend Holm's definition of θ by substituting $s = \sigma + j\omega$ for $j\omega$ which is justifiable by assuming e^{st} rather than $e^{j\omega t}$ as a solution.

(3) Using two-port theory, write the desired transfer function as combinations of these transcendental functions and of the source and load impedances.

(4) The output transform is determined by multiplying the transfer function by the input transform. It is now necessary to transform to the time domain.

(5) To obtain the inverse transform, find the poles of the output transform, which are the roots of a transcendental equation, by numerical or graphical evaluation. There are an infinite number of these poles.

(6) Compute the residue at each pole by application of L'Hospital's rule to the output transform.

(7) By Cauchy's residue theorem, the output function of time is the sum of the residues.

Since there are an infinite number of poles associated with the output transform, the resulting time function is in the form of an infinite series. To yield results of any use such a series must be truncated and the response graphically constructed. An example of this method showing the manipulations and difficulties encountered is given by Gray.[8]

2.3. *Limitations*

Although the inverse transform method, in principle, gives a solution for arbitrary termination and drive, it does not meet the needs for flexible circuit design. The time required to obtain a solution may be reduced by the use of computers, but considerable computational effort is still necessary. Furthermore, the resulting solution does not lend itself to physical interpretation nor does it supply the insight so necessary to the practical circuit designer.

A comparison of the results of this method with the optimum description of transient response for circuit design purposes (Section 1.1) suggests a search for a new approach.

3. The Lumped Model

3.1. *Thin-Film Structure*

Thin-film structures consisting of a dielectric wafer upon which a resistive film has been deposited on one side and an ohmic film on the other (Fig. 1) have been described by Holm.[1] Representative values of dimensions and physical constants are given in Table 2.

Table 2

Representative values for a micro-system circuit

Symbol	Quantity	Representative Value
L	length of film	10^{-2}m
w	width of film	10^{-3}m
k	thickness of film	10^{-7}m
a	thickness of wafer	10^{-3}m
ε	dielectric constant	10^{-8} farad/m
σ	conductivity of film	10^{-3} mhos-□

A 'zero-order' approximation to the time response of the unloaded structure is obtained from the dimensions and constants by the use of familiar relations:

$$R = \frac{1}{\sigma}\frac{L}{kw}, \qquad C = \varepsilon\frac{Lw}{a}$$

and, therefore,

$$RC = \frac{\varepsilon L^2}{\sigma ka}$$

3.2. *A Distributed Model*

The technique of evaluating the transient response by use of the inverse Laplace transform is based on a distributed model of the structure. Such a model, emphasizing differential behaviour, leads to partial differential equations (Table 1) and corresponding transcendental solutions (Section 2.2).

The essential features of this model are

(*a*) At each point two types of current occur:

diffusion current across a finite area A

$$I_{\text{diff}}(x,t) = -\sigma A \frac{\partial V(x,t)}{\partial x}$$

storage current or time rate increase of charge stored in differential volume $A\mathrm{d}x$:

$$I(x,t)\,\mathrm{d}x = \frac{\partial Q(x,t)}{\partial t}\mathrm{d}x = C\,\mathrm{d}x\frac{\partial V(x,t)}{\partial t}$$

(*b*) A continuity relation holds at each point: net current into the differential volume equals the time rate increase of charge stored, that is, the storage current. Hence

$$-\mathrm{d}[I_{\text{diff}}(x,t)] = C\,\mathrm{d}x\frac{\partial V(x,t)}{\partial t}$$

(*c*) At each point, the diffusion equation, derived from relations above, applies:

$$\frac{\partial^2 V(x,t)}{\partial x^2} = \frac{\tau}{L^2}\frac{\partial V(x,t)}{\partial t}$$

The distributed model is an idealization of the actual structure. It neglects:

(*a*) contact resistance;

(*b*) stray capacitances and inductances;

(*c*) inhomogeneity in the materials;

(*d*) finite resistance of the conductive film;

(*e*) surface phenomena.

3.3. *The Purpose of Models*

The model of a device or a process is an idealization that lends itself to mathematical analysis. In the construction of a model one generally emphasizes first-order and neglects second-order effects. For example, to make the problem of heat flow in a rod mathematically tractable, one demands that the model of the process assumes:

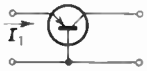
(*a*) the material is homogeneous;

(*b*) the temperature is the same at any given cross section;

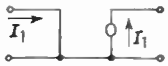
(*c*) the thermal conductivity is not a function of temperature;

(*d*) radiation may be neglected.

With the development of a model, attention shifts from actual device behaviour to model behaviour. The correspondence between actual performance and model performance depends, of course, on the approximations made in deriving the model.



(*a*) Circuit symbol.



(*b*) First-order model.

Fig. 2.

In each engineering application, specific requirements determine the desired degree of correspondence. A seemingly gross idealization is often chosen by design engineers; to illustrate this the first order model for a grounded-base transistor amplifier is examined. Such a model (Fig. 2) emphasizes the essential features—low input, high output impedance and almost unity current transfer—without giving detail which is frequently unnecessary and undesirable in the initial stages of design.

3.4. *Properties of the Lumped Model*

Since the distributed model and its associated inverse transform method is unsatisfactory for circuit design, a change of model is suggested. In the new model (Fig. 3) the voltage, instead of being a continuous variable, is defined at only three points—the two terminals and the middle of the model.

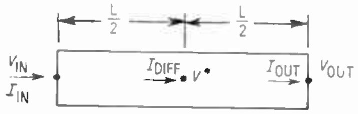


Fig. 3. Definition of variables for one-lump model.

Comparisons of the essential features of the two models given in Table 3 show that:

(*a*) Partial differential equations are exchanged for ordinary differential equations.

(*b*) Since only diffusion current is allowed at the end nodes, all charge storage occurs at the middle node.

(*c*) Both models store the same charge at d.c. as the physical structure and have an equivalent parallel plate capacitance of $\varepsilon wL/a$.

(*d*) Both models have the same d.c. resistance as the physical structure namely $L/(\sigma kw)$.

(*e*) Both models, as well as the actual structure, are symmetrical with respect to input and output.

3.5. *Equivalent Circuit*

An electrical circuit equivalent to the lumped model is given in Fig. 4. Future reference to the lumped model will apply to its equivalent circuit, since a simple electric network is more easily dealt with than

Table 3

Comparison between lumped and distributed models

Type of model	Distributed	Lumped		
Where applicable	at each point	at input node	at middle node	at output node
Diffusion current: $I_{\text{diff}}(x, t)$	$-\sigma A \dfrac{\partial V(x,t)}{\partial x}$	$-\dfrac{\sigma A}{L/2}(V - V_{\text{in}})$	$\dfrac{-\sigma A}{L/2}(V - V_{\text{in}})$	$\dfrac{-\sigma A(V_{\text{out}} - V)}{L/2}$
Storage current: $I(x, t)\mathrm{d}x$	$C\mathrm{d}x \dfrac{\partial V(x,t)}{\partial t}$	none	$\dfrac{\varepsilon LW}{a} \dfrac{\mathrm{d}V}{\mathrm{d}t}$	none
Continuity relations:	$-\partial \dfrac{I_{\text{diff}}(x,t)}{\partial x} = C \dfrac{\partial V(x,t)}{\partial t}$	$I_{\text{source}} = I_{\text{in}}$	$I_{\text{out}} - I_{\text{in}} = -C \dfrac{\mathrm{d}V}{\mathrm{d}t}$	$I_{\text{out}} = I_{\text{load}}$

a set of laws. Note that an analogous equivalent circuit for the distributed model requires an infinite number of elements.

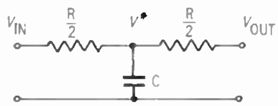


Fig. 4. One-lump equivalent circuit.

This equivalent circuit could have been derived by a different approach. The distributed model corresponds to two-port parameters described by hyperbolic functions (Section 2.2), which may be approximated by finite power series (Table 4). The first terms of the series correspond to the lumped model discussed above. Higher-order lumped models correspond to several terms of the power series.

3.6. *Higher-order Models*

The first-order approximation derived above (Section 3.4) and its equivalent circuit (Section 3.5) will be shown useful for a wide variety of circuit applications. The limitations of the simple model become noticeable if:

(*a*) the driving function and circuit environment emphasize those aspects of the actual structure which are least accurately represented by the model (Example 2, Section 3.7).

Table 4

Power series justification of lumped model

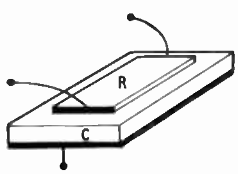
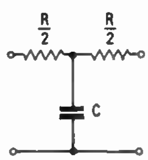
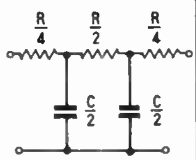
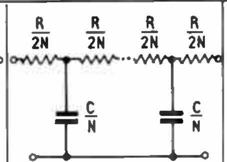
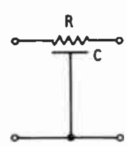
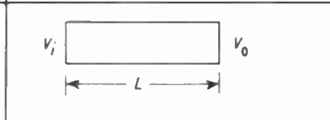
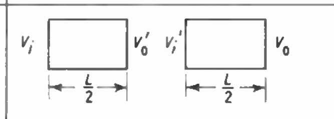
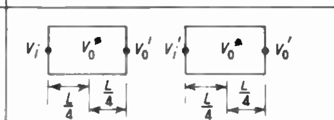
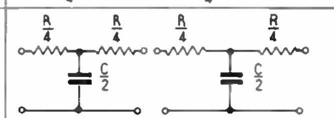
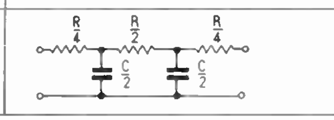
$\theta = (sRC)^{\frac{1}{2}}$, $s = \sigma + \omega$				
				
EQUIVALENT CIRCUIT:	1 LUMP	2 LUMPS	*N* LUMPS	DISTRIBUTED
SUCCESSIVE APPROXIMATIONS:	1st APPROXIMATION	2nd APPROXIMATION	*N*th APPROXIMATION	EXACT
$Ry_{11} = Ry_{22}$	$\dfrac{1+\theta^2/2}{1+\theta^2/4}$	$\dfrac{1+\theta^2/2+\theta^4/32}{1+3\theta^2/16+\theta^4/128}$	$\dfrac{1+\theta^2/2+\theta^4/24+\ldots}{1+\theta^2/6+\theta^4/120+\ldots}$	$\dfrac{\cosh\theta}{\theta^{-1}\sinh\theta}$
$Ry_{12} = Ry_{21}$	$\dfrac{1}{1+\theta^2/4}$	$\dfrac{1}{1+3\theta^2/16+\theta^4/128}$	$\dfrac{1}{1+\theta^2/6+\theta^4/120}$	$\dfrac{1}{\theta^{-1}\sinh\theta}$

Table 5
Construction of a two-lump model

	REPRESENTATION	OPERATION
(1) ACTUAL STRUCTURE		DEFINE TERMINAL VARIABLES
(2) TWO-LUMP MODELS DEFINED		BREAK STRUCTURE INTO TWO LOOPS. DEFINE TERMINAL VARIABLES OF EACH LUMP
(3) VARIABLES OF TWO-LUMP MODEL DEFINED		DEFINE ALL PERTINENT VARIABLES TREATING EACH LUMP SEPARATELY
(4) EQUIVALENT CIRCUIT OF TWO LUMPS		DRAW EQUIVALENT CIRCUIT TO EACH LUMP
(5) TWO-LUMP CIRCUIT		CASCADE NETWORKS TO CONSERVE CONTINUITY

(*b*) the network has a frequency response which cannot be approximated in general shape by the type of lumped circuit selected.

(*c*) accuracy in numerical design calculation is required which exceeds the inherent accuracy of the model. Examples to be discussed give accuracies of about 10%.

For cases in which the limitations of the one-lump model are important, a judicious choice of a higher-order model is needed. Table 5 illustrates an example, the derivation of a two-lump model that usually will give sufficient accuracy for use in decoupling applications (capacitor output). Note that this equivalent circuit has two natural frequencies, thus allowing description of a response which passes through a maximum.

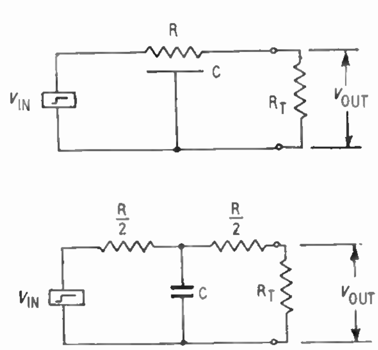


Fig. 5. First-order approximation for common capacitor configuration.

3.7. *Effect of Circuit Environment*

Two examples will illustrate the advantage of appropriately selecting the lumps of a model:

Example 1: The response to a step of voltage is required for the configuration of Fig. 5. The simple model is expected to give fairly accurate response time. Although all the charge is stored in the middle of the model, this will tend to distort only the short time (high frequency) behaviour for the following reasons:

The device is essentially a low-pass filter.

The Fourier spectrum of a step emphasizes the low frequencies.

The simple approximation to the storage effect will therefore not be important and the use of the simple model of Fig. 5 is justified.

Example 2: The response to an impulse of voltage is to be evaluated for the configuration of Fig. 6. The simple one-lump model is useful in the initial stages of design. The long time (low frequency) behaviour will be faithfully portrayed by this model since it is equivalent to the quite accurate, distributed model for d.c. excitation.

On the other hand, the short time response will not be given accurately for the following reasons:

The actual device emphasizes high frequencies.

The Fourier spectrum of an impulse does not emphasize low frequencies.

In the model, current is passed to the middle and

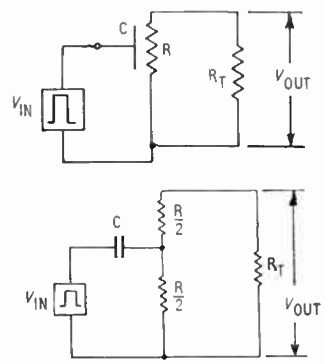


Fig. 6. First-order approximation for capacitor input configuration.

then 'diffuses' to the ends; in the actual device, there is much more direct 'feed through' to the load.

To improve the short time behaviour, lumps are chosen to give the model shown in Fig. 7, which emphasizes the capacitance 'feed through' to the ends and still maintains a good deal of simplicity. This corresponds to choosing lumps as in Table 5, but defining the voltage not in the middle of each lump, but rather one-quarter of a lump length from the terminals. Choice of the model is somewhat arbitrary. In practice, a good choice strongly depends on how familiar the designer is with the nature and requirements of the specific problem.

3.8. *Effect on Waveshape*

The simple model has one storage element and hence one natural frequency associated with it, although the overall circuit may contribute natural frequencies as well. Hence, in evaluating the step response for the common capacitor configuration with resistance load (Fig. 5), the one-lump (or one-pole) model will give a simple exponential rise.

Since it is evident that the voltage must diffuse across the actual device, an initial condition at the output is $dv/dt = 0$. This fact is not disclosed by the model. Frequently, however, quite accurate waveforms can be sketched on a linear time scale by starting the output voltage waveform with zero slope and then superimposing the exponential given by the simple model. Higher order models give a better approximation to the zero initial slope, since the second and

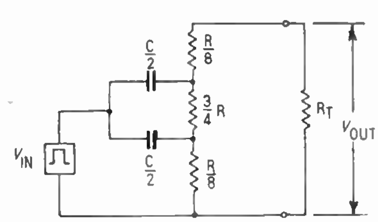


Fig. 7. Second-order approximation for capacitor input configuration.

higher order natural frequencies will tend to match the slope as well as the amplitude of the response with increasing accuracy.

The delay time due to diffusion in the distributed structure was found experimentally for several structures. A good empirical 'rule of thumb' for the magnitude of delay time appears to be:

$$\text{delay time} \simeq RC/10$$

Chow[10] derived an analytical expression for delay resulting from the finite time needed for carriers to travel across the base of a diffusion transistor. Since processes in the micro-system network and the transistor base are analogous if recombination is neglected, expressions for micro-system delay can be obtained from Chow's work by setting the reciprocal of the minority-carrier lifetime equal to zero. This is not done here because the delay will not be significant in most design problems. An ideal delay line (Fig. 8) can be added to the one-lump model to give a circuit which usually quite accurately describes all the essential features of the output waveform.

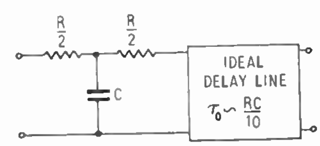


Fig. 8. First-order approximation corrected for high-frequency transients.

3.9. *Justification of the Lumped Model*

The essential difference between the lumped model and inverse transform approach is in the 'timing' of approximations. Starting with the actual structure, both methods neglect secondary effects and arrive at a set of laws describing the phenomena of prime importance. Here the attacks diverge. The inverse transform method aims at the solution of the partial differential equations to yield the desired output transform in terms of combinations of transcendental functions. A numerical or graphical evaluation of the poles of the output transform allows the use of the Cauchy residue theorem to give a solution in the form of an infinite series. The series is truncated and the solution constructed graphically. Note that an approximation is made in the last step of analysis in order to get a usable result.

In the lumped model approach, the approximation is made much earlier to avoid working with partial differential equations and their transcendental solutions. The distributed character of the structure is ignored by treating the phenomena on a finite rather than differential basis. This lumping leads to ordinary rather than partial differential equations and to a simple, and hence practical, equivalent circuit which

Table 7

One-lump T-equivalent circuit

(*a*) *Voltage Drive*

R_T/R \ C_T/C	large	small	absent
large	1/2	1/2	1/2
small	C_T/C	1/2	1/4
short	C_T/C	1/2	1/4

(*b*) *Current Drive*

R_T/R \ C_T/C	large	small	short
large	R_T/R	$(R_TC_T)/(RC)$	R_T/R
small	1/2	$C_T/(2C)$	1/2
short	1/2	1/2	1/2

ther apart, the capacitor approaches the usual physical notion of an open circuit. However, pushing the plates closer together gives a structure which is capable of charge storage, if the two plates do not touch. Since we may only approach infinity, never quite getting there, the almost-infinite capacitor has one property that a short circuit does not have, charge storage. Hence, although the correspondence of an almost-infinite capacitor and a short circuit is valid for steady state behaviour since neither may have a voltage developed across it then, such a concept is misleading for evaluation of transient behaviour. The large capacitor differs from a 'short' in its transient response radically since it takes a long time for a charge to be collected in such a structure.

From the equivalent circuit, the response time for a purely resistive load is

$$\tau_c = \frac{R_T+R/2}{2RC(R_T+R)}$$

4.4. *Response Time for an Arbitrary Load*

Formulae for the response time of an arbitrarily loaded circuit will be useful to:

(*a*) derive asymptotic behaviour;

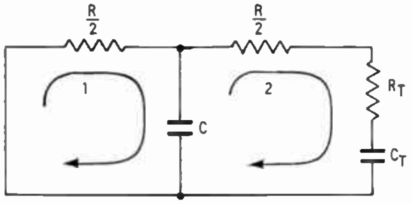


Fig. 12. Loop currents for transients of circuit in Fig. 11.

(*b*) obtain points for construction of the design chart, particularly those points outside the region of validity of the asymptotes.

Setting the determinant of the impedance matrix of the circuit of Fig. 12 to zero

$$\begin{bmatrix} R/2+(pC)^{-1} & -(pC)^{-1} \\ -(pC)^{-1} & R/2+R_T+(pC)^{-1}+(pC_T)^{-1} \end{bmatrix} = 0$$

yields the characteristic frequencies,

$$p_{1,2} = (-1 \pm \sqrt{1-\Delta})/(2\tau_s) \quad \text{where } \Delta = 4\tau_s/\tau_L$$

and

$$\tau_L = [(C_T/C)(R_T/R+1)+\tfrac{1}{2}]RC$$

$$\tau_S = [(C_T/2C)(R_T/R+\tfrac{1}{2})/\tau_L](RC)^2$$

Now Δ has a relative maximum for $C_TR_T = RC/2$ and with this constraint an absolute maximum for small values of C_T. If C_T exceeds $C/2$, or if C_TR_T differs from $RC/2$ significantly, then Δ is much smaller than unity. For example $\Delta_{max} = 0{\cdot}67$ for $R_T = R$ and $C_T = C/2$ and $\tau_S/\tau_L = 0{\cdot}17$.

Thus it is generally permissible to set

$$\sqrt{1-\Delta} = 1-\Delta/2$$

with less than 15% error even for the above value of Δ_{max} and with negligible error for all asymptotic cases. Two approximate roots result

$$p_1 = -1/\tau_L \qquad p_2 = -1/\tau_S$$

which are the negatives of the reciprocal time constants.

Asymptotic limits of τ_L and τ_S give three distinct values for large or small values of R_T/R and C_T/C as plotted in Fig. 13(*a*). The three regions are bounded by zones in which the asymptotic formulae do not necessarily hold.

4.5. *Evaluation of Amplitudes*

For the circuit of Fig. 10, the response to a voltage step input is

$$v(t) = 1 - L\exp(-t/\tau_L) - S\exp(-t/\tau_S)$$

since $v(\infty) = 1$

Further

$$v(0) = 0 \quad \text{yields } L+S = 1$$

and
$$\frac{dv(0)}{dt} = \frac{4R_T}{RC(2R_T+R)} = \frac{L}{\tau_L}+\frac{S}{\tau_S}$$

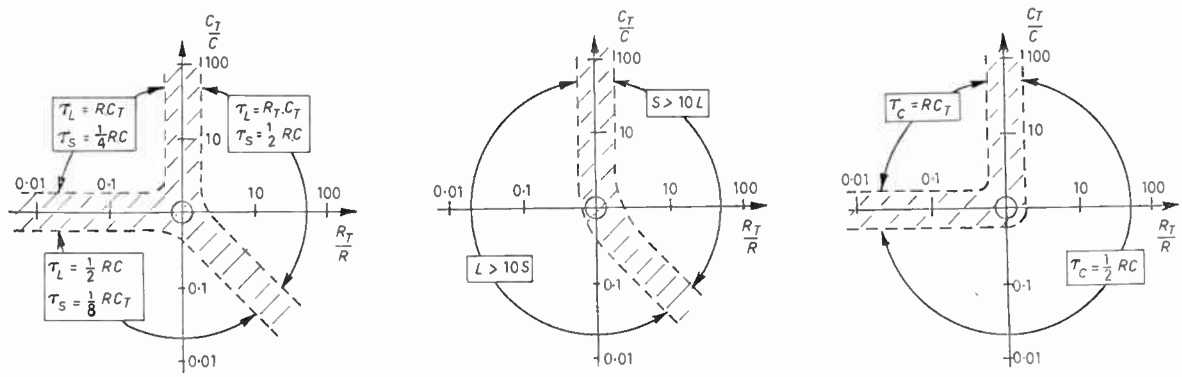
Calculation of L/S as a function of R_T/R and C_T/C gives the two asymptotic regions, shown in Fig. 13(*b*), namely

$$L > 10S \quad \text{and } S > 10L$$

as well as a boundary region for intermediate values.

4.6. *Asymptotes*

The time constants τ_L and τ_S and the corresponding amplitudes L and S, plotted in Fig. 13(*a*) and (*b*), determine the circuit response τ_c shown in Fig. 13(*c*).



(*a*) Time-constants. (*b*) Amplitudes. (*c*) Dominant time-constants.

Fig. 13. Asymptotic approximations for transient response.

Clearly, if L is large, τ_c will approach τ_L, conversely if S is large, τ_c will approach τ_S; therefore, two regions result in Fig. 13(*c*) which are separated by a boundary zone and which correspond to values given in Tables 6 and 7 for asymptotic approximations. The resulting asymptotes form the skeleton for construction of the design charts (Fig. 14).

4.7. *Design Chart*

Values not obtainable from asymptotic approximation are in the regions $R \sim R_T$ and $C \sim C_T$ shown as shaded zones in Fig. 13(*c*). The design chart Fig. 14 was derived by numerical point-by-point calculation as follows:

(*a*) compute τ_L and τ_S;

(*b*) compute L and S;

(*c*) compute $v_0(t)$;

(*d*) compute τ_c such that $v(\tau_c) = 0{\cdot}63$.

5. Experimental Results

5.1. *Experimental Device*

Methods of fabrication were those reported by Happ, Fuller, and Castro[11] and Wagner and Happ.[12] The dimensions of the experimental device and its essential characteristics are shown in Table 8.

Table 8

Experimental device

R	5·7 kΩ	Measured on impedance bridge
C	4200 pF	Measured on impedance bridge
$\tau = RC$	24 μs	Time constant
ε	1500	Barium titanate compound
L	0·0254 m	Film length
w	0·0064 m	Film width
k	$1{\cdot}1 \times 10^{-5}$ m	Resistive film thickness
σ	70 mhos/m	Film conductivity
a	$5{\cdot}1 \times 10^{-4}$ m	Substrate thickness

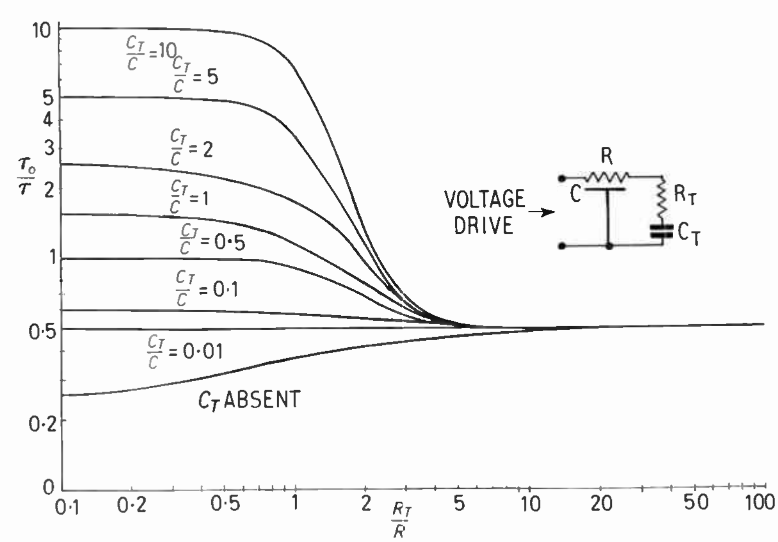


Fig. 14. Design chart calculated from one-lump model.

5.2. *Instrumentation*

The open-circuit transient response of thin-film structure with small time-constants cannot be accurately measured with commercially available oscilloscopes. The input impedance of common oscilloscopes is 10 to 50 pF in parallel, with 1 to 10 MΩ, giving an impedance in the order of only 10 kΩ at 1 Mc/s. This poor approximation of an open circuit distorts the short-time response.

To make the measurements as accurate as possible, a cathode follower was used in the circuit of Fig. 15 to provide a high output impedance to the distributed-parameter network. The impedance of the cathode follower, experimentally determined, is a 30 MΩ resistor in parallel with a 3 pF capacitor.

5.3. *Response Time for Resistance Load*

Response times for the micro-system terminated in pure resistance appears in Table 9 with the corresponding calculated values, and is plotted in Fig. 16.

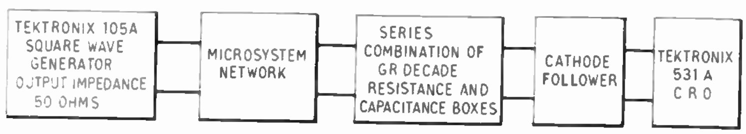
5.4. *Response for Arbitrary Load*

The calculated and measured response times of the micro-system terminated in a series combination of resistance and capacitance are compared graphically in Fig. 17. Calculated response times are taken from the design chart (Fig. 14.)

5.5. *Waveforms*

For large terminating capacitance, the magnitude of the load resistance has a profound effect on the

Fig. 15. Instrumentation.



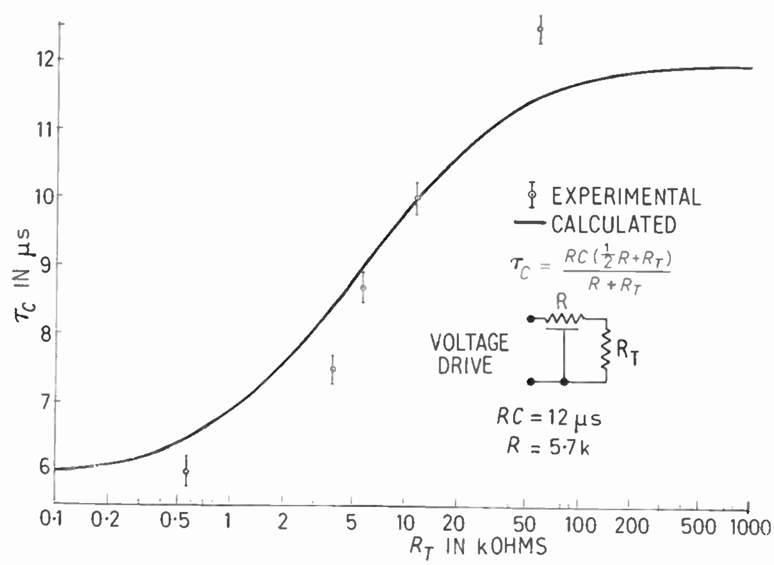


Fig. 16.
Response time against resistance load.

Table 9
Comparison of calculated and measured response times for resistive load

R_T kilohms	τ_c calculated microseconds	τ_c measured microseconds	Waveform	Remarks
0·57	6·5	6·0	(*a*) Experimentally, waveform follows closely for all R_T as predicted	10 to 20% experimental accuracy
3·82	8·4	7·5		
5·7	9·0	8·7	(*b*) Approximately 2 μs time delay	
11·4	10·0	10·0		
57	11·5	12·5		

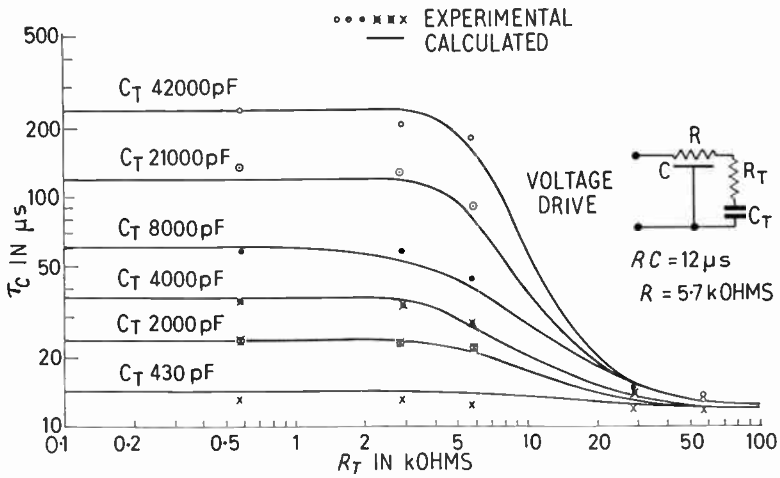


Fig. 17. Experimental and calculated responses.

response wave shape; for small load capacitance, the effect is smaller. This is illustrated in Fig. 18. Note also the comparison between calculated and experimental curves.

6. Typical Applications

To this point the lumped model has been driven only by a perfect voltage source and terminated in series resistance and capacitance; however, the model may be quite generally used in any circuit environment. A few examples will illustrate other typical applications.

6.1. *Finite Source Impedance*

The response $v_0(t)$, calculated from both a one- and a two-lump model, is compared in Table 10. The

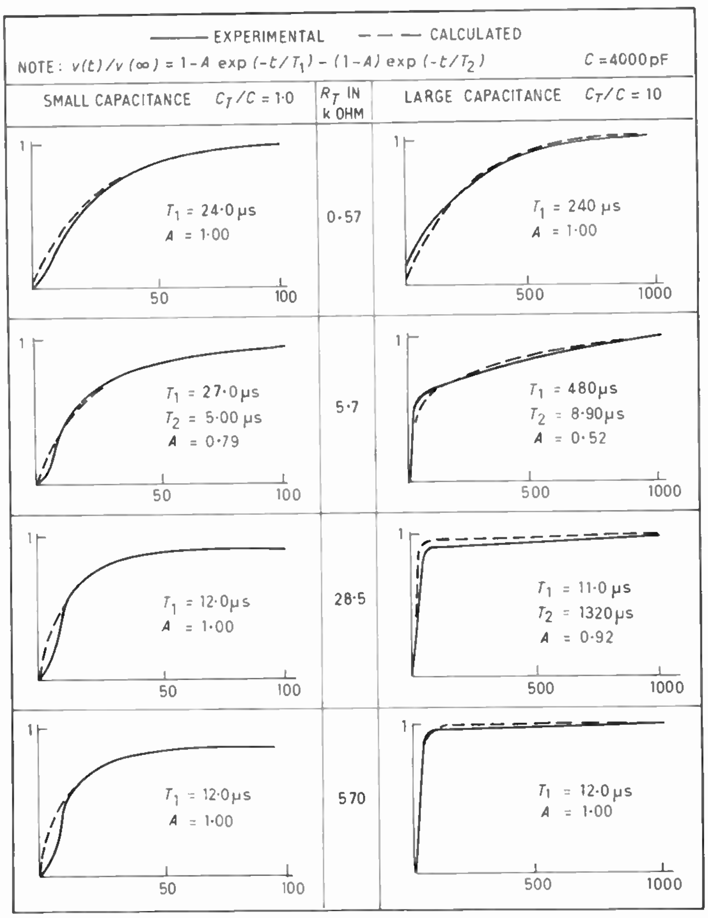
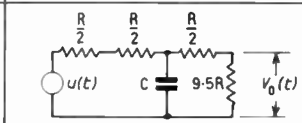
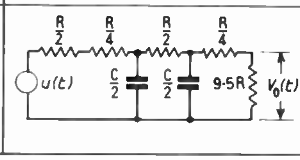


Fig. 18. Waveforms.

Table 10
Comparison between models

MODEL	EQUIVALENT CIRCUIT	$V_0(t)$	$\frac{T_1}{\tau}$
ONE-LUMP		$0{\cdot}860(1-\exp-t/T)$ $T = 0{\cdot}91\tau$	0·91
TWO-LUMP		$0{\cdot}860-0{\cdot}99\exp-t/T_1$ $-0{\cdot}125\exp-t/T_2$ $T_1 = 0{\cdot}82\tau$ $T_2 = 0{\cdot}10\tau$	0·90

expressions for $v_0(t)$ were derived by use of standard techniques of network analysis (Fig. 19).

6.2. *Waveforms*

The network tested in Section 5 was used in the circuit of Fig. 20 to determine the response experimentally. In Fig. 21 the calculated and experimental responses are compared. Note the improvement in the one-lump approximation in short time response by inclusion of a delay line in the model (Fig. 8).

6.3. *Other Orientations*

As with a transistor, or for that matter any three-terminal device, there are three distinct orientations

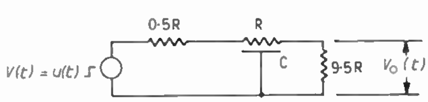


Fig. 19. Finite source impedance.

of the micro-system structure. These are listed in Table 11. In contrast to the common capacitor, the capacitor input and capacitor output circuits exhibit high-pass characteristics; hence, we expect that the one-lump model will not be as accurate (see Sect. 3.6) for these configurations as for the low-pass common capacitor orientation.

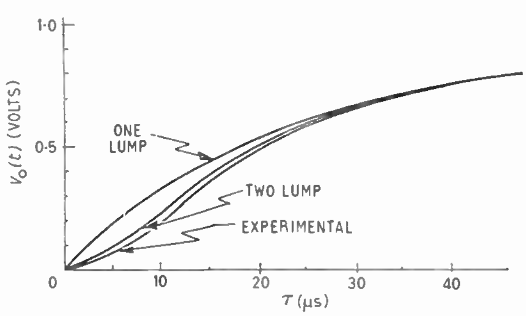
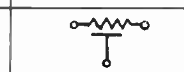
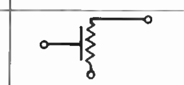
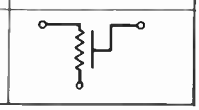


Fig. 21. Comparison of calculated and experimental response.

Table 11
Network configurations

CONFIGURATION	CODE	STRUCTURE
COMMON CAPACITOR	CC	
CAPACITOR INPUT	CI	
CAPACITOR OUTPUT	CO	

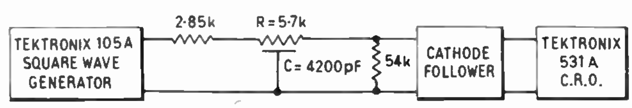


Fig. 20. Circuit to determine waveform.

6.4. *Capacitor Input*

Consider an open circuit micro-system in the capacitor input orientation driven by a unit step from a voltage source (Fig. 22). The one-lump approximation gives

$$v_0(t) = \exp -(2t/\tau)$$

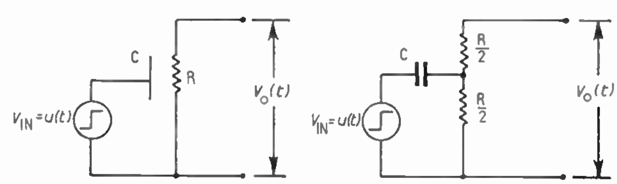


Fig. 22. Capacitor-input configuration.

The device of Section 5.1 was used in the test set-up of Section 5.2 to determine the response experimentally. This response, which appeared to be purely exponential, had a time constant of 10·5 μs which compares favourably with the calculated response time of 12 μs.

The previous example should not imply that the one-lump model for capacitive input will give 15% accuracy for arbitrary load and source impedances; in fact, this seems highly unlikely (Section 3.6).

A comprehensive study of model behaviour for the capacitor input and capacitor output orientation aiming at the prediction of accuracy for an N-lump model in arbitrary circuit environment would be of practical importance.

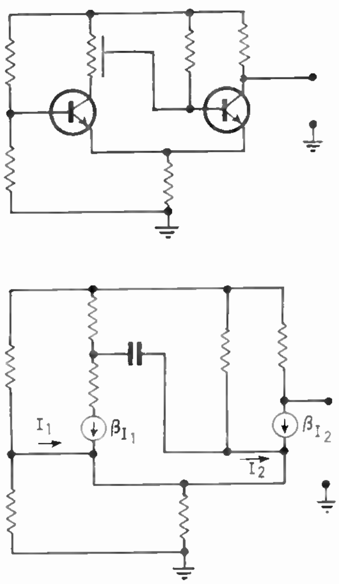


Fig. 23. Emitter-coupled multivibrator.

6.5. *Microminiature Emitter-coupled Multivibrator*

To study the design feasibility of an emitter-coupled multivibrator consisting of two transistors and thin films deposited on a single dielectric wafer, the following lumped-parameter model (Fig. 23) is useful: The thin-film structures are replaced by one-lump models and the transistors are replaced by current generator models. These steps are necessary since distributed models for either the transistors or the thin-film structures lead to an intractable mathematical problem. The model presented permits evaluation for monostable operation of this multivibrator.

7. References

1. C. Holm, "Distributed parameter networks for microminiaturization", Radar Research Establishment Technical Note No. 654, and *Electrical Manufacturing*, pp. 92–6, April 1960.
2. P. S. Castro and W. W. Happ, "Distributed parameter circuits and microsystem electronics", *Proc. Natl Electronics Conf.*, **16**, pp. 448–60, 1960.
3. W. W. Happ and T. R. Nisbet, "Using Jacobians for frequency-selective networks", AIEE Pac. Gen. Conf., Paper CP60–1128, San Diego, also *Electronic Ind.*, **20**, pp. 86–9, January 1961.

   T. R. Nisbet and W. W. Happ, "The calculus of deviations applied to transistor circuit and network analysis", *J. Brit. I.R.E.*, **21**, pp. 437–50, May 1961.

   W. D. Fuller and W. W. Happ, "Design procedure for film-type distributed parameter circuits", *Proc. Natl Electronics Conf.*, **17**, pp. 597–610, 1961.
4. W. W. Happ and P. S. Castro, "Distributed parameter circuit design techniques", *Proc. Natl Electronics Conf.*, **17**, pp. 44–70, 1961.

   P. S. Castro and W. W. Happ, "Subnetworks", *I.R.E. Intl Conv. Rec.*, **10**, Part 2, pp. 3–7, 1962.
5. J. Crank, "The Mathematics of Diffusion" (Clarendon Press, Oxford 1956).
6. H. S. Carlslaw and J. C. Jaeger, "Conduction of Heat in Solids", 2nd ed. (Clarendon Press, Oxford, 1959).
7. J. G. Linvill, "Lumped models of transistors and diodes", *Proc. Inst. Radio Engrs*, **46**, pp. 1141–52, June 1958.
8. P. E. Gray, "The Dynamic Behaviour of Thermoelectric Devices (Technology Press of M.I.T. and John Wiley, New York 1960).
9. D. L. Starner, "NOR Speedup with a Distributed Parameter Network", Tech. Report LMSD 895021 (Internal communication as referenced and summarized by Castro and Happ[2]).
10. W. F. Chow and J. J. Suran, "Transient analysis of junction transistor amplifiers", *Electronics*, **26**, No. 11, pp. 189–91, November 1953, and *Proc. Inst. Radio Engrs*, **41**, No. 9, pp. 1125–29, September 1953.
11. P. S. Castro, W. W. Happ and W. D. Fuller, "Synthesis of solid state distributed parameter functions", *Inst. Radio Engrs Conv. Rec.*, **10**, Pt. 6, pp. 262–78, 1962.

    W. W. Happ, "Synthesis of distributed-parameter *RC* networks", *Proc. Inst. Radio Engrs*, **50**, No. 483-4, April 1962.

    W. W. Happ, "Stray capacitance in thin-film circuit", *Electronic Industries*, **21**, No. 6, pp. E4–E7, June 1962 and *U.S. Govt. Research Report*, **37**, No. 8, p. 18(A), April 20 1962.
12. W. W. Happ and O. E. Wagner "Field emission of rutile diodes", *J. Appl. Phys.*, **34**, No. 2, February 1963.

Manuscript first received by the Institution on 16th November 1962 and in final form on 8th July 1963. (Paper No. 865.)

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".

AN ISOCHRONOUS CYCLOTRON

An isochronous cyclotron under construction in Holland will enable different kinds of ions to be accelerated to a variable final energy (protons up to a maximum of 25 MeV). In an isochronous cyclotron the influence of the relativistic increase of mass on the revolution frequency is compensated by making the induction $\bar{B}(r)$ of the magnetic field increase with the radius r. The vertical stability of the orbits is ensured by making B vary periodically in the azimuthal direction by means of sector-shaped hills and valleys on the pole pieces. The Philips cyclotron is also fitted with ten pairs of trimming coils, which allow the form of $\bar{B}(r)$ to be varied to meet the demands imposed by the different ionic species and the different final energies. This article describes how this complicated field is measured—for ten different excitations of the main coil, in connection with the desired variation of the final energy—and how the more than 10^5 measured values are processed with a computer. The final result of the calculations is the value of the currents through the trimming coil needed to give the best approximation to the desired isochronous field. In the course of these calculations, improbable measured values are indicated and replaced.

"Investigation of the magnetic field of an isochronous cyclotron", N. F. Verster and H. L. Hagedoorn. *Philips Technical Review*, **24**, pp. 106–20, 1962/63.

INTERFERENCE IN TELEVISION RECEPTION

The type of interference which may produce picture jitter in fly-wheel synchronized domestic television-receivers is explained in a recent German paper. The appropriate interference characteristics, i.e. the $\phi - \phi$ characteristic for additive interference and the $\Delta\phi - \omega$ characteristic for interference from f.m. modulated sync pulses, are discussed and compared with each other. In doing so it emerges that picture jitter due to line frequency fluctuations increases when the noise bandwidth is reduced, i.e. when the circuit becomes less sensitive to additive interference. The relationship between these two types of interference is also determined by the filter in the control loop. Suitable design data are given for the usual *RC* filter. With the aid of the relationship derived in this way it is possible to calculate fully the fly-wheel synchronization under conditions of frequency modulation of line frequency and other synchronization properties.

"The sensitivity against interference of television receiver circuits with fly-wheel synchronization", A. Grünewald. *Nachrichtentechnische Zeitschrift*, **16**, pp. 368–78, 1963.

V.H.F. PROPAGATION IN MOUNTAINOUS REGIONS

When v.h.f. radio-links in mountainous regions are being evaluated, reflection and attenuation due to obstacles are to be taken into account. For reception in the shade of a mountain-ridge the transmission attenuation is always at least 6 dB greater than the free-space attenuation. Calculation shows that if propagation is to take place over great distances, the presence of a mountain-ridge between transmitter and receiver can yield gain as compared with the case when there is a spherical, smooth earth's surface between them. The reflection and diffraction of radio waves and the calculations of the transmission loss of two given routes are given in a paper by a Dutch engineer. The measurements appear to give reasonable conformity with the theory. However, an accurate forecast is impossible because of the many reflections occurring in practice. It remains essential that the most favourable position of the receiving and the transmitting antenna should be determined by means of field-strength measurements.

"The propagation of radio waves in mountainous regions", J. W. A. v.d. Scheer. *Tijdschrift van het Nederlands Radiogenootschap*, **27**, No. 6, pp. 287–95, 1962.

INTERCHANNEL INTERFERENCE IN MULTI-CHANNEL TRANSMISSION

Experiments have been carried out in Japan to reduce interchannel interference by using differential gain equalization when transmitting up to 1800 telephone channels by microwave frequency modulation. Interchannel interference on a comparatively small number of multiplex telephone channels is mainly caused by delay distortion of the transmission system. However, as the number of multiplex telephone channels becomes much larger (over a thousand), it is found that the effect of amplitude frequency characteristics of the f.m. carrier circuit cannot be ignored. The analysis of the differential gain characteristics produced by the f.m. transmission circuit shows that its inclination and curvature correspond well with the second and third order inter-modulation caused by amplitude-frequency characteristics. By equalizing the differential gain characteristics, using an experimental transversal equalizer, an improvement of the signal-to-interchannel interference ratio has been obtained. The improvement of signal-to-noise ratio amounted to about 8 dB.

"Improvement of interchannel interference by using differential gain equalizer", M. Kuwabara and T. Matsumoto. *Journal of the Institute of Electrical Communication Engineers of Japan* **45**, No. 10, pp. 1319–24, October 1962.