THE USE OF ELECTRONIC TECHNIQUES IN THE CONSTRUCTION OF MACHINES WHICH CAN LEARN.

AWARDED DEGREE OF M.Sc.

BY JOHN FREDERICK YOUNG

SUMMARY

Four different approaches to the design and construction of electronic machines capable of displaying Pavlovian conditioning by association have been investigated theoretically and practically. Conflicts in the basic requirements for such a machine might be resolved by the introduction of Probability, Inhibition and Forgetting. The methods adopted should not make it impossible to introduce these features as and when required later.

The methods of association information storage considered have been :-

1. Use of recordings of sinusoidal signals resulting from beats between sinusoidal input signals. This introduced problems of linearity in the recording process, and of selection of the input frequencies.
2. Use of rectangular waves of various frequencies. While this method could eliminate linearity problems, it introduced problems of addition of new information to previously-recorded information.
3. Use of electrical pulse signals recirculating dynamically. While use of this method could eliminate many of the earlier problems, the problem remained of noise augmentation in a recirculating system. It is desirable to reduce this if possible.
4. Use of pulse signals recirculating non-dynamically.

This fourth method proved to be the most promising, and a machine was constructed using this approach. The organisation of this machine was based on the earlier work. Methods for further simplification are suggested in the thesis.

Work on this fourth method involved the following :-

1. Consideration of methods of avoidance of spurious response and of possible alternative methods of storage of information.

2. Development of various forms of prime-number counter. The method finally adopted used integrated circuits.

3. Investigation of a number of circuits displaying majority logic action.

4. Relationship to other work using multi-dimensional theory, and the simplification of the visualisation of associations.

5. Theoretical investigation of magnetic pulse recording waveshapes.

6. Development of methods of construction of the machine.

CONTENTS

1. INTRODUCTION

1. INTRODUCTION

The capabilities of man-made control systems have been increased spectacularly in recent years. The performance of early servo systems was limited to some extent by the fixed nature of their control characteristics. During the past decade, adaptive control systems have been investigated intensively. In such systems the control characteristics are altered automatically so as to maintain a desired standard of performance despite variations of system parameters. Analogies can be drawn between adaptive action in control systems and homeostatic behaviour in living creatures, such a comparison being one function of cybernetic studies.

The writer has recently carried out a thorough review of this field, as a preliminary to the present research. This review has been published as the textbook "Cybernetics" [1], which is submitted in support of this thesis. Consequently, no general review will be given in the present thesis, but the general background will be taken as known. However, the particular background on which the present research has been based will be discussed.

The investigation is based on the fundamental problem of how humans and animals recognise speech sounds and assign meanings to the sounds. The eventual object is to devise a machine which might be used to carry out such recognition, in the hope that the difficulties encountered will give some clue to the elucidation of the fundamental problem. We can recognise, for example, the word "cat" spoken by people with many different accents and even if the speech waveforms are degraded in various ways. It is not clear exactly which features of the speech waveforms are used for the generalisation required in such a recognition process. It is hoped that investigation of the type of machine envisaged in the present research might eventually lead to an identification of these features. Animals and humans are inadequate subjects for use in such research, since by their nature they are capable of adapting to compensate for various forms of degradation of information. Indeed, the very nature of this adaptation is of great interest, but it is difficult to investigate it since it cannot be controlled in the animal and the human.

High-speed electronic digital computers are now widely used, and methods of programming them to carry out an extensive range of mathematical operations have been devised. Digital computers have been used to simulate many forms of biological activity, including some of the actions of the nervous system. In biological terms, a digital computer is a purely reflex device which will, one hopes, always infallibly carry out a particular action in response to a given stimulus. This of itself does not limit the use of digital computers in the simulation of animal nervous systems, but when combined with high cost and limited capacity and availability it helps to restrict the range of such applications. In these circumstances, the design and construction of special-purpose cybernetic machines is an attractive alternative to the use of programmed general-purpose computers. The present research is an investigation of possible methods of construction of one such machine.

It is sometimes suggested that a permanent form of information storage simulating the function of memory is not essential for research purposes. However, research should make possible repeatable experiments, and unless some permanence can be ensured if and when required in the memory process, repeatable experiments are not possible. The method adopted should not therefore make permanence of memory impossible.

In work such as that discussed here, it is necessary to consider carefully whether an attempt is to be made to duplicate biological action precisely or whether it is only the overall effect which is to be copied.

Perhaps the best illustration of the applicability of this approach is the development of heavier-than-air flight. The man-made aeroplane was developed not as a result of mere slavish copying of biological mechanisms. Instead, a study of biological principles, together with the application of the best engineering techniques available at the time, led to the development of the flying machine. While the principles employed depended to a great extent on what had been learned from the study of living creatures, the machine was in no way a simple copy.

Modern techniques of integrated microcircuit production make possible the construction of cybernetic electronic devices which in some cases are approaching in size the smallness of the equivalent biological nervous system. At the same time, investigations of biological nervous systems are making clearer some of the ways in which these systems are organised. Because of these facts, the use of electronic devices to simulate some aspects of biological nervous action is continually becoming easier.

2. PREVIOUS WORK

2. PREVIOUS WORK

It is perhaps as well to consider at this point the relationship between the present work and some of the other cybernetics work which has been reported. The intention of the present work is to design and construct machines in which the facility for learning by association is a built-in feature. Such a machine will be capable of associating without the aid of a human teacher. Much of the work which has been reported in this field relies on a deliberate reinforcing process by a human teacher, who adjusts the machine in accordance with its responses. It is sometimes suggested that feedback is not a necessary feature of a learning process.[2] However, without any form of feedback, changes of environment will introduce ambiguity in the actions of a machine. The ambiguity can be made less important by the introduction of a forgetting process, but inevitably the machine can not then be completely error-free. Entirely random inputs to a non-feedback machine are meaningless and cannot be used in the learning process, since in time every input would be associated with every other input. Feedback need not necessarily be deliberately built into a machine, since it can occur via the environment.
In a feedback system, the effects of the active system outputs are detected and used to provide additional inputs.

In recent years there has been extensive investigation in many parts of the world into numerous forms of trainable learning machine.[3] When examined closely, much of the latter work is seen to belong to a single class. This might be called the "Adjustable-Weight, Majority Logic" approach. In this, each of a number of sensors, such as photo-electric cells, provides an input which is taken, via an adjustable "weighting" resistor individual to that input, to a single common point, the input terminal of an amplifier. A given fixed procedure or strategy is then followed in order to adjust the weighting resistors in such a way as to ensure that the amplifier gives a particular output only when a given set of inputs is applied. In this way, the machine is gradually trained, or adjusted, to recognise patterns. The training process is carried out in general by a human operator. Convergence to a final "trained" state has been obtained by following certain training strategies, proved theoretically to lead to this convergence.[4]

In general, the aim of such work has been the eventual production of working pattern recognition machines, rather than the simulation of animal actions. This is a good engineering object, since such machines are required for use with electronic computers, for example to enable computers to operate on information produced on normal typewriters.

The writer has pointed out elsewhere that such a system is not well adapted to operation under changing conditions once the training period is over.[5]

In animal systems there is no definable training period followed by the achievment of a trained state. Such a system would fail to adapt to changing conditions and such failure could possibly be fatal to the organism. A single severe error is sufficient to produce fatality. Consequently, such non-adaptive action would tend to be eliminated by the operation of natural selection processes.

This fact alone would seem to explain why the animal learning process never produces complete freedom from error. Although the probability of error is reduced with the length of the training period, the necessary facility for adaptation ensures that error is never completely absent. If learning machines are to have a facility for adaptation, then it would appear to be desirable that the "training phase" never ends, that the machines operate on a probabilistic basis, and that therefore the operation will never be completely error-free even under fixed input conditions. There are many human examples of illusions caused by this effect. Perhaps the outstanding example is the illusory change in a spoken word reiterated over and over again from a tape recorder. In such a fundamental respect, the machines envisaged here will differ from most machines which have been investigated.

3. CONDITIONED REFLEX ACTION

3. CONDITIONED REFLEX ACTION

In the seventeenth century, Descartes discussed the occurrence in animals of reflex activity.[6] Early in the present century, I.P. Pavlov investigated conditioning of reflex action in animals.[7] In his classic series of experiments he showed that if a bell was sounded each time that food was presented to an animal, eventually the bell alone could cause salivation, even though the food was not then presented. It is possible to explain much animal learning action in terms of conditioning of reflexes.

In effect, one stimulus "A" is associated with another stimulus "B" repeatedly until eventually stimulus "A" can on its own produce the response normal to stimulus "B". It is important to note that the resulting conditioned response is not infallible. It appears rather that the probability of producing response "B" by stimulus "A" is increased by the repetition of the association. Since it is possible to explain a wide range of animal learning activity on this simple basis, it is of engineering interest to consider the feasibility of machines having similar capabilities.

Simple machines in which one or two inputs can be associated have been constructed and widely demonstrated.[8] However, it is not easy to extend such simple machines to handle a wide range of different inputs.[9] The present research is aimed at the construction of a machine which can be extended, once the basic principles are clarified.

The means adopted for information storage simulating the function of memory in some simple machines have produced traces which were transitory in the extreme. Examples are the storage of voltage on capacitors[10] and decaying oscillation in electronic circuits[11]. Such methods are not adequate if a thorough investigation of an extended machine is envisaged. Some of the methods of information storage used in digital computers are much more permanent, but are too inflexible for initial use in the research. An example is core storage.[12] The means adopted should preferably give permanent storage regardless of temporary power supply failure. The storage form should give "non-destructive read-out", i.e. use of the stored information should not cause its destruction. The storage of new information should not affect previously stored information. However, there should be provision for the inhibition of older conditioned reflexes. The provision of some form of slow forgetting mechanism at a later stage must be kept in mind.

The overall effect of a machine based on the requirements above should be probabilistic. It is not required that, once stimulus signals S_a and S_b have coincided, stimulus S_a should always inevitably produce effect e_b corresponding to stimulus S_b. Rather it is required that the probability of production of effect e_b by stimulus S_a should increase with the past frequency of occurrence of the coincidence (S_a, S_b).

There are thus three requirements for a store which can be used to display Pavlovian conditioning:-

1. Small size, preferably molecular.
2. Exclusive recording of coincidence.
3. Probabilistic recording of coincidence.

An additional desirable factor would be that older recordings should decay slowly, so that they are of less importance than are more recent recordings. This "forgetting" process appears to be an important feature of animal activity.[13] While it is not essential to incorporate a forgetting process into initial work, the eventual need must be kept in mind. In effect, such a process helps to avoid overloading of the memory. In the animal case, it enables the animal to discard habits which are no longer of use in a changed environment or at a more advanced age. It is also desirable that an inhibitory process can be introduced if required.

4. BASIC REQUIREMENTS

4. BASIC REQUIREMENTS

The initial basic requirements of an associatory learning machine are as follows :-

1. It must have some means of detecting coincidence between the occurrences of a number of input stimuli..
2. It must have some means of recording these coincidences.
3. It must have some means of making use of a recorded coincidence if any of the input stimuli occur in the future.
4. It should preferably operate on a probabilistic basis.

There have recently been many different approaches to the construction of learning machines. The following are a few of the disadvantages noted by the writer in the course of reviewing such work.[1]

1. Difficulty of extension. Whatever practical approach is adopted, it should not be limited to use with only a few inputs. Although it is difficult at present to envisage the extension of man-made devices to a biological level of complexity, the possiblity of extension should be as unlimited as is practicable. This implies that sub-units should be simple, small and inexpensive.

2. Error-free learning. Freedom from error has been widely used as a criterion of excellence in, for example, pattern-recognition machines.[14] However, such a criterion implies that each error must have a constant definition. This is not true of biological systems, which in general possess the property of adaptation to slow changes of the environment. The possession of the facility for "forgetting" is implied in this.

3. Two-dimensional approach. It is a surprising fact that many workers have ignored the three-dimensional nature of the nervous system. This is particularly true of investigations of networks of artificial neurons, which have usually been limited to the two-dimensional approach.[15] (In some cases these networks have been simulated on a digital computer.[16])

In an associatory learning machine, if input signals S_a and S_b occur simultaneously, then a record $R(S_a,S_b)$ must be made. This record must be kept available so that if signal S_a occurs alone at a future time, then the effect e_b of signal S_b can be produced even though signal S_b is not actually occurring.

The recorded signal should be characteristic only of the associated input signals. For example, if the input signals are $S_a,S_b,S_c,S_d........S_z$, then it must be true that the recorded signals :-

$$R(S_a,....S_b) \neq R(S_x,.....S_y)$$

where $S_x,....S_y$ are any signals other than $S_a,....S_b$. It should be noted that this relationship must be fulfilled even if, to take a simple two-member example, $x = a$ but $y \neq b$. In this case again, we must have :-

$$R(S_a,S_b) \neq R(S_a,S_y) .$$

Thus any particular recorded coincidence $R(S_a,....S_b)$ must be exclusive to a particular set of signals $S_a,....S_b$ and must never correspond to any other set, including any set which differs from the original set by only one member.

It is an additional requirement that each and every subset coincidence should be separately detected and recorded, since each of these provides information which might be required by the machine at some future time. Indeed in the case of an organism, subset coincidence might be of vital future interest. Consequently it is necessary to arrange that :-

$$R(S_a,S_b,S_c,S_d,\ldots S_z) = R(S_a,S_b) + R(S_a,S_c) + \ldots \ldots + R(S_a,S_z) + R(S_b,S_c) + R(S_b,S_d) + \ldots R(S_b,S_z) + R(S_c,S_d) + \ldots R(S_c,S_z) + \ldots \ldots \ldots + R(S_y,S_z)$$

It is **supposed** that associations are recorded in pairs. This will include associations in threes, fours, etc., provided that there is a high probability that each and every pair association is recorded. One effect of this requirement is that the device is then capable of pattern completion or of operating on incomplete patterns of input stimulation.

An example will help to illustrate the requirements and difficulties. Suppose that there are four possible input signals S_a, S_b, S_c, S_d, and that the following associations occur :-

$$R(S_a,S_b), \quad R(S_a,S_c), \quad , R(S_a,S_c,S_d)$$

Now if all subset pairs are separately recorded, then

$$R(S_a,S_c,S_d) = R(S_a,S_c) + R(S_a,S_d) + R(S_c,S_d)$$

If this condition is met, then at some point in the future input signal S_a appearing without input signal S_c will be capable of producing association S_d, so that wrong information is stored. Signal S_a should not produce signal S_d unless signal S_c is also present. Thus it is necessary to ensure that

$$R(S_a,S_b,S_c,\ldots) \neq R(S_a,S_b) + R(S_a,S_c) + \ldots$$

It should be noted, however, that if this requirement is rigidly enforced then the machine will inherently only be capable of a minimal amount of pattern-completion.

It can be seen that there appear to be conflicting requirements :-

1. All subset pairs should be separately recorded, since every pair can provide useful information for pattern-completion.

2. Subset pairs should not be separately recorded, since it is possible for a pair to provide incorrect information.

In the animal, it is possible that the conflict is resolved by three features :-

1. Probability
2. Inhibition
3. Forgetting.

If the first of these features is incorporated, then the strength of the recorded information will depend on the frequency of past occurrence of the particular association. Consequently, information will be available for pattern completion. However, those associations which occur only coincidentally with other associations will be recorded less strongly than those associations which occur additionally in isolation.

The second feature would involve the addition of inhibition of associations which have not occurred in the past. While such a process possibly occurs to some extent biologically, it is not practicable to incorporate inhibition of every non-coincident association, since this would require excessive storage capacity.

The third feature is known to occur biologically. Its use would ensure that memory of those associations which occurred only in coincidence with other associations would decay more rapidly than the more frequent associations which also occurred in isolation.

Since the best method to resolve the conflict in an associatory learning machine is not known at this stage, it is desirable that all three features should be capable where necessary of separate introduction, modification and elimination in the machines envisaged here. It will then be possible to carry out comparative tests on the use of the three features.

5. CONDITIONAL PROBABILITY

5. CONDITIONAL AND JOINT PROBABILITY

It was shown in section 3 that a machine should be probabilistic, i.e. the probability of production of an effect e_b by a stimulus S_a should increase with the past frequency of occurrence of the coincidences of S_a with S_b.

Consider two events A and B which occur in the sequences shown in fig. 5.1, where each occurrence is marked by an asterisk.

```
   A      ******
   B      *********************
 Time    ----t1--------------t2---
```

fig. 5.1

If the conditional probabilities $P_A(B)$ and $P_B(A)$ are estimated from previous events, then at time t_1 :- $P_A(B) = 1$ and $P_B(A) = 1$, while at time t_2 :- $P_A(B) = 6/21$ and $P_B(A) = 1$.

It is sometimes suggested that since at time t_2 :-

$$P_A(B) \neq P_B(A)$$

then these two probabilities should be recorded separately.[17] No doubt from a simple and purely mathematical point of view this is correct, but such a view ignores the physiological phenomenon of "forgetting". If the point t_2 is far enough removed in time from point t_1, then the early association between A and B would be almost completely forgotten by any animal system.

We have no information on the exact law of forgetting, but it is tempting to postulate an exponential decay of memory traces.

$$\text{Now at time } t_1 , \quad P_A(B) = P_B(A)$$

$$\text{but at time } t_2 , \quad P_A(B) < P_B(A)$$

Thus any scheme involving the use of conditional probabilities must incorporate separate storage facilities for $P_A(B)$ and $P_B(A)$.

To obviate such separate storage facilities, and to give an improved economy of storage, it is suggested that the memory system ought simply to store P(A&B). Then at time t_1 above, P(A&B) = 1

but at time t_2 above, P(A&B) = 6/21,

i.e. P(A&B) has decayed because of non-occurrence of the coincidence. Thus it is proposed that storage facilities are provided not for Conditional (or Unidirectional) Probabilities, but rather for Joint (or Bi-directional) Probabilities.

If there are N inputs A,B,C,.....N,

then the Conditional or Unidirectional Probabilities are

$P_A(B)$, $P_A(C)$, $P_A(D)$,..........$P_A(N)$,

$P_B(A)$, $P_B(C)$, $P_B(D)$,..........$P_B(N)$,

$P_C(A)$,.........etc.

There are therefore N(N-1) Unidirectional probabilities to be stored, if two-way associations are recorded.

However, the Bidirectional Probabilities are :-

P(A&B),P(A&C),P(A&D),...............P(A&N),

P(B&C),P(B&D),...............P(B&N),

P(C&D),... etc.

There are therefore $\frac{N(N-1)}{2}$ of these Bidirectional Probabilities for which storage provision must be made. The storage capacity ratio between Bidirectional and Unidirectional methods is therefore one-half, if two-way associations are recorded.

6. GENERAL THEORETICAL BACKGROUND

6.1 LEARNING AND FORGETTING CURVES

If it is assumed that the quantity to be stored is P(A&B), then it is possible to draw learning and forgetting curves for different circumstances by plotting the value of P(A&B) against time. This is done in fig.6.1 It is assumed here that the stimulus B is repeated continuously while stimulus A appears as a short train of stimuli. The shape of the resulting probability curve depends on the instant of occurrence and the length of pulse train A. Transfer from a learning curve to a forgetting curve takes place at the end of the A pulse train.

It should be noted that the relationship between the probabilities $P_A(B)$ and P(A&B) is given by :-

$$P_A(B) = \frac{P(A\&B)}{P(A)}$$

Now the quantity P(A) is always fractional or equal to unity, so it follows that :-

$$P_A(B) \geqslant P(A\&B)$$

The curves of fig.6.2illustrate the relationships for a practical case.

It will be desirable to introduce such probability relationships into the operation of the machines envisaged here. While it is not absolutely necessary to incorporate these arrangements at the commencement of the work, the methods adopted must not prevent such incorporation at a later stage.

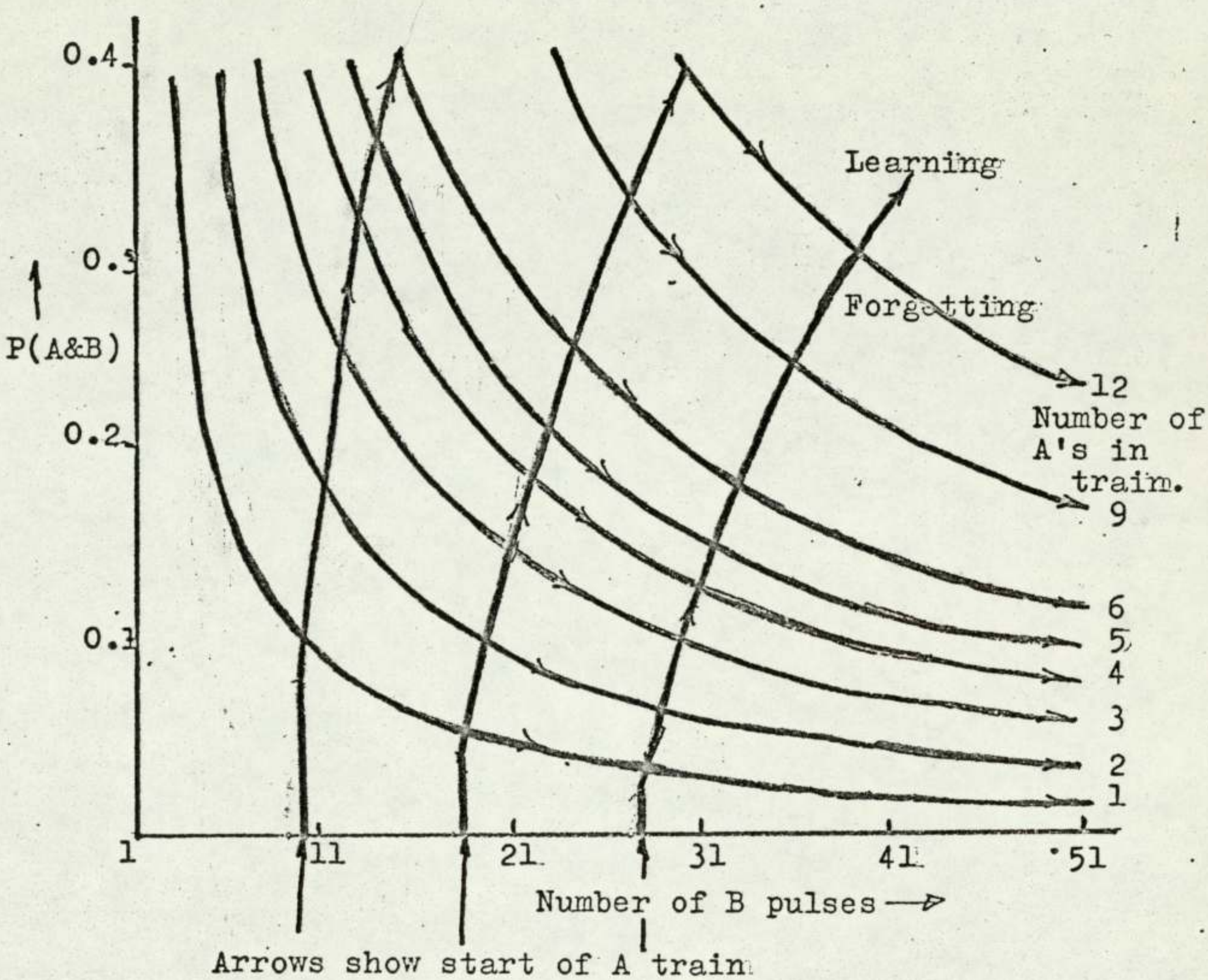


fig. 6.1

LEARNING AND FORGETTING CURVES

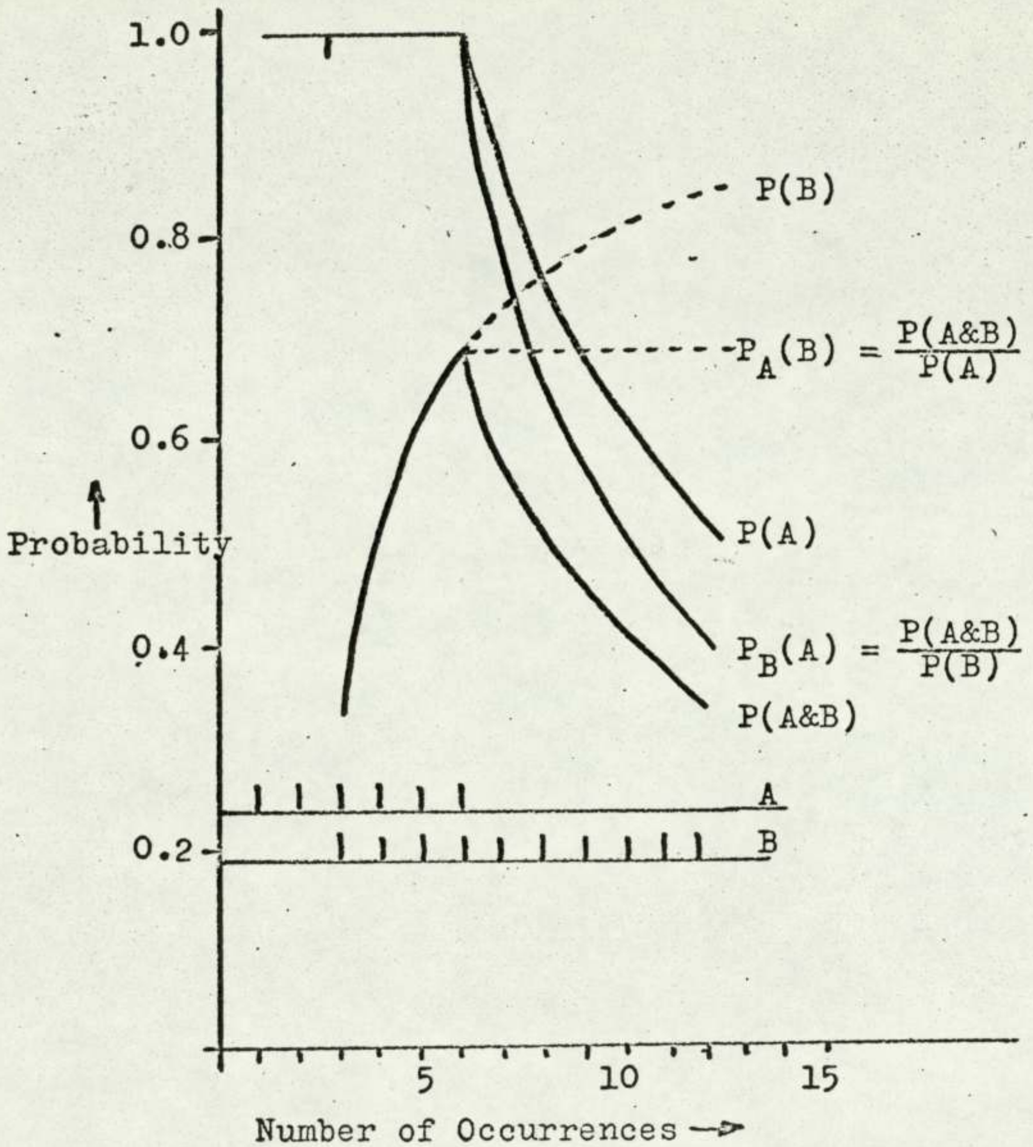


fig. 6.2

PROBABILITY RELATIONSHIPS FOR A PARTICULAR CASE

6.2 REPRESENTATION OF HYPERSPACE

It is useful to be able to express association relationships geometrically, one reason being the ease of visualisation of the results. The representation of associations by the use of hyper-space has been widely used. In this approach[18], the points representing different input patterns are regarded as being separated by hyperplanes which are implemented in practice by simple majority logic circuits. Unfortunately it is impossible for a human to visualise a hyperspace directly, and published representations[52] have merely used a simple two-dimensional drawing of a cube in 3-space for illustration, as below in fig. 6.3a.

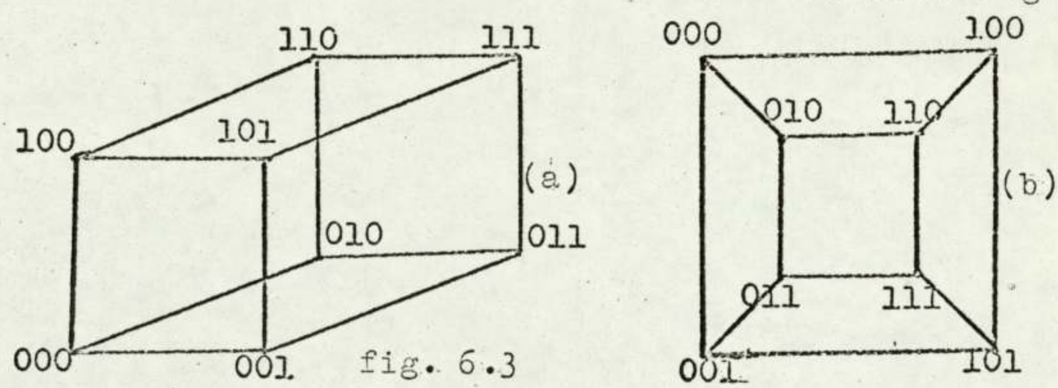


fig. 6.3

It would be possible to make use of the well-known representation of a hollow hypercube for visualisation, though this fact does not seem to have been exploited. A hollow cube with binary numbered vertices can be represented in two dimensions as shown above in fig. 6.3b. In a similar way, a "hollow" four-dimension hypercube can be represented in a two-dimensional drawing as in fig. 6.4. Thus it is quite possible, even though complex, to represent hypercubes by means of two-dimensional sketches.

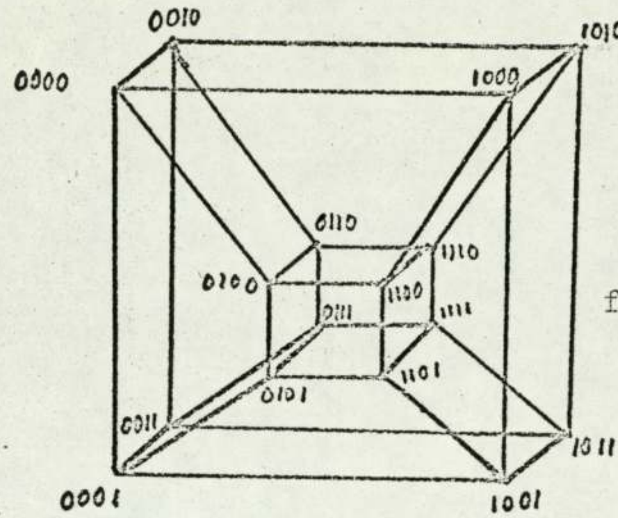


fig. 6.4

A useful alternative two-dimensional sketch of a 3-space representation of the associations between three inputs is shown below. This gives a more logical layout of the vertices than do those of fig. 6.3.

In the first representation of fig. 6.3a, the point 000 can be taken as the origin and each of the three coordinate axes is then taken as representing one of the input variables, which can each take the value 1 or 0. In the new representation of fig. 6.5 the diagram has been rotated and the layout of vertices has been changed slightly in order to bring all numbers containing the same number of binary ones to a common level.

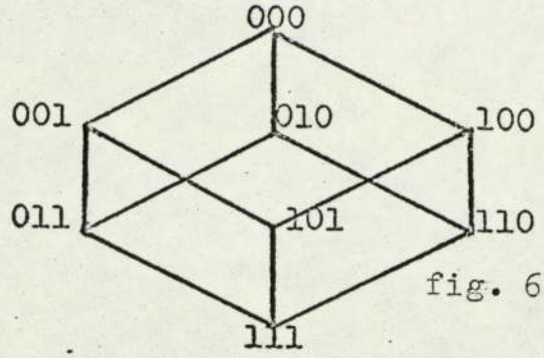


fig. 6.5

By extending this representation, the four-dimensional hypercube representing all possible multi-way associations of four inputs can be drawn in two dimensions as shown below. Compared with the previous representation, the inner cube is now "suspended" below the outer cube.

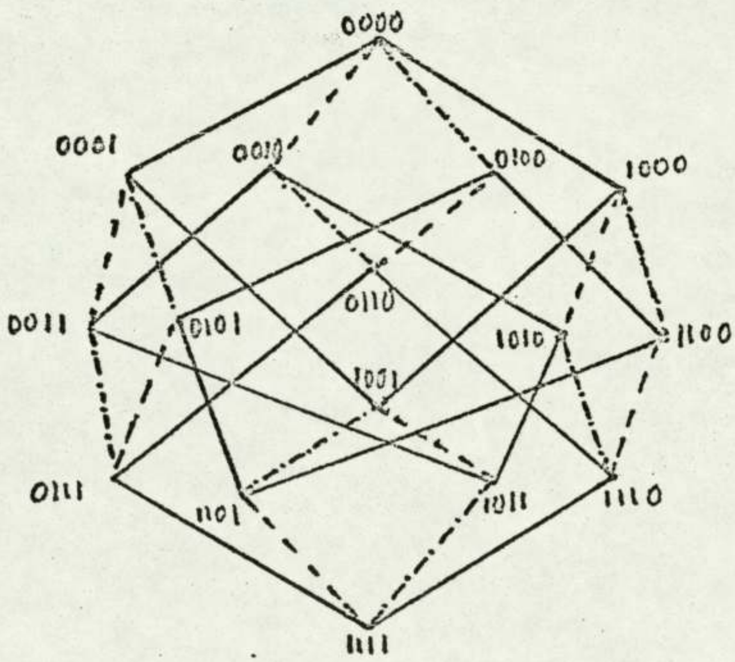


fig. 6.6 Representation of Hypercube.

It will be noticed that the numbers of vertices in the different horizontal arrays are equal to the Pascal binomial coefficients, for example, 1,4,6,4,1. The total number of vertices is equal to the sum of all of the Pascal coefficients, which is in turn produced by a binary progression. For example, $1+4+6+4+1 = 16 = 2^4$. The number of connections going downwards from each vertex decreases uniformly as the diagram is traversed from top to bottom, for example 4,3,2,1 in the diagram shown.

Four edges terminate at each vertex. The binary numbers of the vertices joined by any single line differ by only one digit. Every plane is common to two 3-cubes and on the diagram each 3-cube is simply represented by twelvelines. Some of the 3-cube representations are very distorted. An example has vertices :-

0001-0000-1000-1001; 0011-0010-1010-1011.

Another has vertices :-

0011-1011-1111-0111; 0010-1010-1110-0110.

Thus it is quite possible to represent a hypercube by a two-dimensional drawing and to use this to study the connectivities between vertices. As the number of dimensions of the hypercube increases, so the inter-connections of the diagram become extremely complicated, but the layout of the vertices in two dimensions is quite straight forward.

To separate out a number of hypercube vertices for recognition, all lines on the diagram joining any required vertices to any non-required vertices must be cut by the separating hyperplanes.[19] In order for it to be possible to separate out a set of points using only a single hyperplane, it must be possible to draw along the lines of the diagram which pass through all points of the required set a joining line which does not pass through any other points which are not in the set. The separating hyperplane must then pass through all of those other lines of the diagram which connect points of the required set to points which are non-members of the set. If it is not possible to draw such an exclusive joining line, then more than one hyperplane is required for the separation. As a simple illustration of this, two separate intersecting planes would be required to obtain the function A(Bc + bC) in three dimensions, i.e. to separate out the points 101 and 110. On the other hand, only one plane is required for the function A(B+C), which involves the separation out of the points 101, 110, and 111.

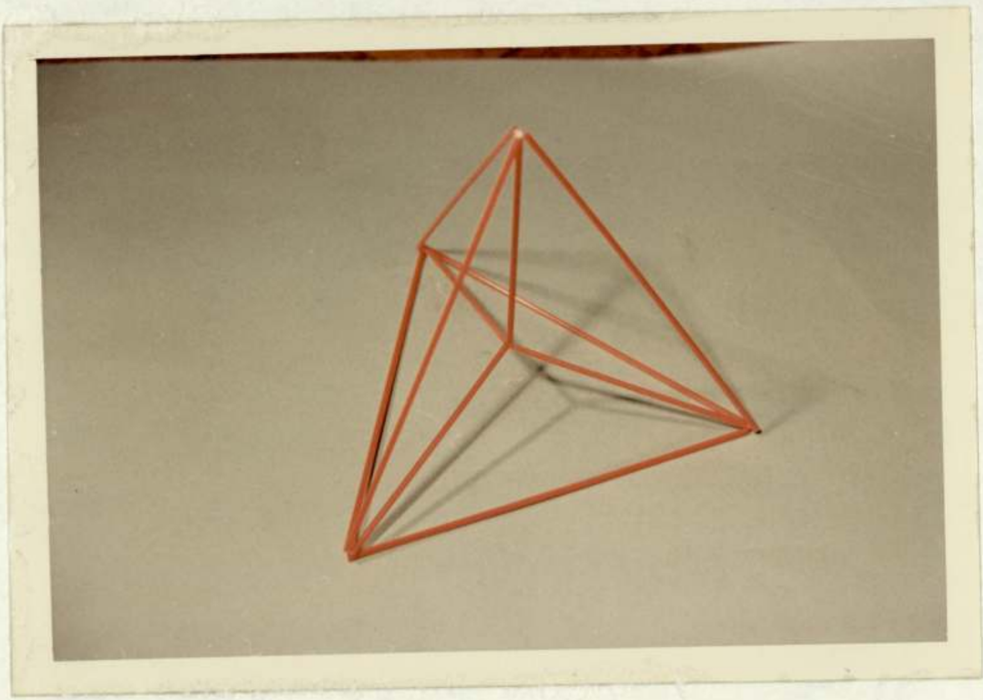
It might be convenient to study the surface formed by joining the mid-points of those lines which connect points which must be separated, e.g. the mid-point of the line joining vertex 101 to vertex 100 in the last example. If necessary, the nature of this surface could be investigated by comparison with the various planes joining the vertices. The effect of a threshold change in a majority logic gate is to move the corresponding hyperplane vertically, while independent changes of individual coefficients rotate the hyperplane.

6.3 GEOMETRIC REPRESENTATION OF ASSOCIATIONS

It is useful to be able to visualise the principle of operation of a machine. In some of the published work on character-recognition, for example, every possible set of conditions of the N inputs has been represented by a different vector from a set in N-dimensional space[30] as described above.

In the present work, consideration is given to associations between pairs of inputs. If each input is represented by a point, then the possible associations between inputs can be represented by lines joining the points. For two inputs, A and B, a single straight line represents the association AB. For three inputs A,B,C, a triangle AB, BC, CA, is required. The six possible two-way associations between four inputs are represented by a pyramid, each side being an equilateral triangle. To represent the ten possible two-way associations between five inputs, a four-dimensional hyper-pyramid is required. A three-dimensional sketch model looking "inside" such a hyperpyramid is shown. On the model, as on any drawing, distances are distorted.

Now if more than five inputs are to be represented, figures in a multi-dimensional space are required. It was realised that any of these could be sketched, with distance-distortion, in three dimensions. This in turn led to a two-dimensional distorted representation.



THREE-DIMENSIONAL REPRESENTATION OF A FOUR-DIMENSIONAL HYPERPYRAMID

It is convenient to have all input points A,B,C... equispaced around a circle, and to represent the associations between inputs by straight lines. The thickness or number of straight lines joining any two input points can then be used to indicate the strength of the association between these points.

This method of representation illustrates clearly that the greater the number of inputs, the more difficult becomes the problem of distinguishing one association from another. The arbitrary assumption of orthogonality often used[3] implies that each situation is completely separable from any other situation.

The real justification for the use of orthogonal representation is that it places every situation next to every other situation. This is also true of the simplified method of representation of two-way associations given here, and the representation can easily be drawn or visualised for any number of inputs.

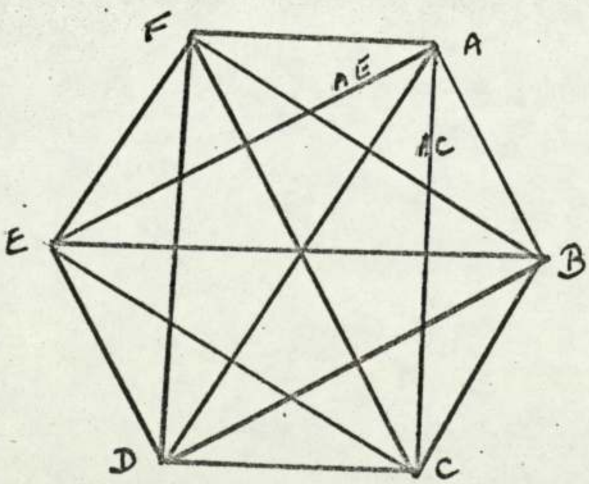


fig. 6.7

REPRESENTATION OF ASSOCIATIONS

7. SINUSOIDAL RECORDING

7.1 USE OF SINUSOIDAL RECORDING

In early attempts to produce an associatory learning machine, the writer made use of the properties of non-linear electrical circuits. For example, if two sinusoidal voltages having different angular frequencies are added together and applied to a squaring element such as a Hall crystal,[20] then only one of the output terms has a low angular frequency. This term can easily be separated out by filtering. The fact that such filtering is necessary if the low-frequency term is to be separated out might make it un-necessary to use expensive squaring devices such as Hall plates and make it possible instead to use any simple non-linearity which will introduce cross-modulation.

Now consider a number of sinusoidal oscillators each producing a different angular frequency w_x whenever it is energised by the occurrence of an input. If the outputs at frequencies w_a and w_b of any two oscillators are applied simultaneously to a non-linear modulator, then a modulator output containing the angular frequencies $2w_a$, $2w_b$, $w_a + w_b$ and $w_a - w_b$ is obtained. The latter, low-frequency term cannot always be filtered out from the rest if the range of input frequencies exceeds one octave. It is therefore necessary to use a single sideband modulator.

The basis of a possible association-recording machine based on the use of this single sideband modulator principle is shown in fig 7.1

If, for example, oscillators a and b are energised simultaneously, producing signals at angular frequencies of w_a and w_b, then a signal of frequency $w_a - w_b$ is recorded on the magnetic tape loop. The signal $w_a - w_b$ can then be picked up from the tape via the replay arrangements at any time in the future to indicate that the two oscillators a and b have been simultaneously energised at some time in the past.

Before considering further how the recorded information might be used, it will be as well to consider obvious difficulties with the system described. There are two main problems :-

1. The recorded informatiom can be ambiguous if the frequencies $w_a \ldots\ldots w_z$ are not carefully chosen. For example, if $w_c - w_b = w_y - w_x$, then there is no way of distinguishing between recorded signals indicating an association of c with b and those indicating an association of x with y. In order to avoid such ambiguity, a careful choice of oscillator frequencies is necessary, and this choice is not easy if a large number of inputs has to be handled.

2. In the system as shown in the diagram, it is

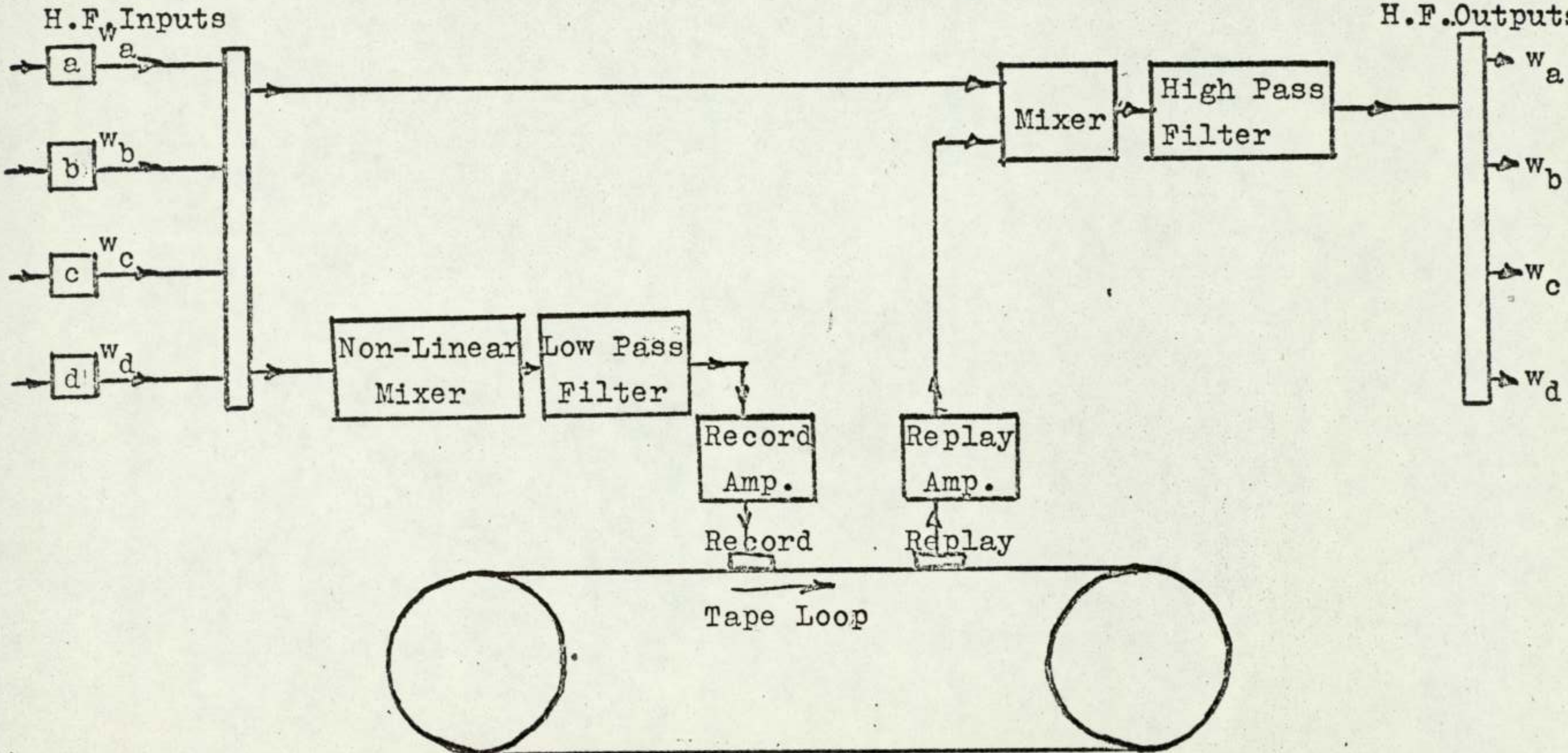


fig. 7.1 BLOCK DIAGRAM OF ASSOCIATION-RECORDING MACHINE

necessary to record newly-acquired information on to the magnetic medium directly over older recorded information, without first erasing the latter. This procedure introduces problems of partial erasure of older information and of loss of information due to tape saturation, and it is therefore not a desirable mode of operation.

In order to avoid the first difficulty, it is necessary to construct tables of numbers having exclusive differences (or possibly exclusive sums). To illustrate the problem, a simple difference table is given below.

f	Differences						
10							
11	1						
13	2	3					
17	4	6	7				
22	5	9	11	12			
30	8	13	17	19	20		
40	10	18	23	27	29	30	
54	14	24	32	37	41	43	44
69	15	29					

The above difference table has been constructed by inserting at the beginning of each line of differences the first difference integer which has not previously been used in the table. Inevitably, eventually one encounters a number which has previously appeared and which is not therefore available. The final entry (29) in the above table is such a case, 29 having been used two lines previously.

In the above example, there are only eight integral values of f which can be used from the 44 equally spaced frequency channels nominally available. There are different ways of constructing such tables. For example, the second line of differences could have started with 3 instead of 2, giving :-

f	Differences						
10							
11	1						
14	3	4					
19	5	8	9				
25	6	11	14	15			
32	7	13	18	21	22		
42	10	17	23	28	31	32	
44	2	12	19	25	30	33	34

Here the difference 2 has been introduced near to the end of the table. Eight channels are now usable from only 34 instead of the 44 channels required by the previous table.

It is of interest to consider the effective channel utilisation in the cases considered. The utilisation can be defined as :-

$$\text{Channel Utilisation} = \frac{\text{Number of Usable Channels}}{\text{Number of Available Channels}}$$

Then for the first case considered above :-

No. of Channels	1	3	7	12	20	30	44
Usable Channels	1	3	6	10	15	21	28
Utilisation %	100	100	86	83	75	70	63 %

while for the second case above :-

No. of Channels	1	4	9	15	22	32	34
Usable Channels	1	3	6	10	15	21	28
Utilisation %	100	75	66	66	68	66	82 %

Consider yet another possible difference table :-

2					
3	5				
4	7	9			
6	10	13	15		
8	14	18	21	23	
11	19	25	29	32	34

For this case :-

No. of Channels	2	5	9	15	23	34
Usable Channels	1	3	6	10	15	21
Utilisation %	50	60	66	66	65	39 %

7.2 DISCUSSION OF DIFFERENCE TABLES

The difference tables given have been constructed empirically, and such a process could be programmed on to a digital computer. However, it was felt that such a fundamental problem of channel selection should be avoided at this stage of the work. One requirement is the eventual extension of the number of available channels at least to twenty and preferably beyond, while limiting the necessary total number of channels. It would also be desirable if possible to limit the rapid increase of the values in the first difference column in order to equalise the spacing of the channels as far as possible. If the original frequencies f could always be equal in value to the logarithms of prime numbers, then they would always have exclusive sums. Unfortunately however, such frequencies would be non-integral. In addition, the channel spacing would decrease as the numbers increased. These facts make the log(prime) approach of little value even though published tables of such numbers exist.[21]

The absolute values of frequencies f would of course be of little importance, since the table of differences is the real basis of construction. In order to use other frequencies, the whole table can be multiplied by a constant, or a constant can be added to the f column.

7.3 CONTINUOUS RE-RECORDING

The recording of newly-acquired information over previously-recorded information introduces difficulties as mentioned earlier. In order to avoid such problems, the more complex arrangement shown in fig.7.2was considered and tried.

Here, the information recorded on the tape is picked up by a replay head. In addition to being taken to the output for external use, the replayed information is added linearly to new inputs and the sum is recorded. An erase head ensures that the tape is magnetically "clean" before it reaches the recording head. With this system, the entire information about the past action of the machine is stored on the short length of tape between the record and replay heads. Such an arrangement can be thought of as being analogous to the neural recirculating loops believed to form part of the short-term memory system in animals.[37]

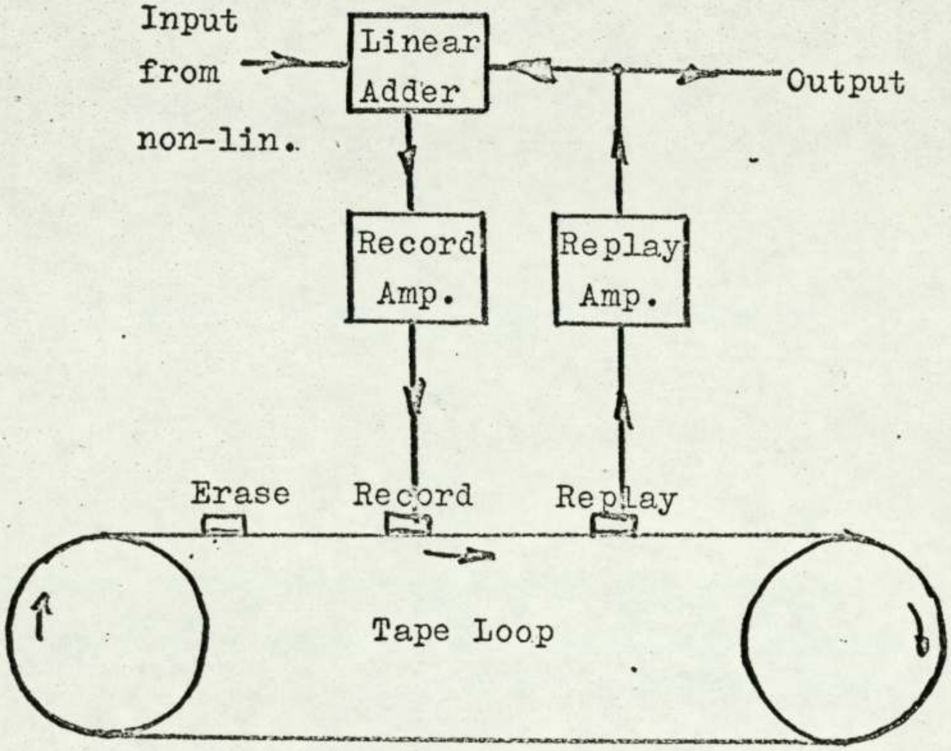


fig. 7.2

BASIS OF MORE COMPLEX ASSOCIATION-RECORDER

7.4 ELECTRONIC RECIRCULATING STORAGE

Recirculating neural loops make possible at least temporary storage of information. The present aim is to produce some of the characteristics of biological systems electronically. Information storage using recirculating loops is not unknown in electronic systems, and it is therefore of interest to study some of the characteristics of such systems.

Consider a positive feedback system having an input signal V_1, a forward transfer function A exp(-sT), an output signal V_o, and a positive feedback function B from output to input. We can write :-

$$V_o = (V_1 + BV_o)A \exp(-sT) \qquad (1)$$

whence :-

$$\frac{V_o}{V_1} = \frac{A \exp(-sT)}{1 - AB \exp(-sT)} \qquad (2)$$

so that :-

$$\frac{V_o}{V_1} = A \exp(-sT) \sum_{n=0}^{n=\infty} A^n B^n \exp(-nsT) \qquad (3)$$

Practical systems have a finite bandwidth, so we can take the upper limit of summation as n = N-1 :-

$$\frac{V_o}{V_1} = A\exp(-sT)\sum_{n=0}^{n=N-1} A^n B^n \exp(-nsT) \qquad (4)$$

Expansion and subsequent contraction of this series of limited bandwidth gives :-

$$\frac{V_o}{V_1} = A\exp(-sT)\left[\frac{1-(AB\exp(-sT))^N}{1-AB\exp(-sT)}\right] \qquad (5)$$

$$= \frac{A\exp(-sT) - A^{(1+N)}B^N\exp(-sT(1+N))}{1-AB\exp(-sT)} \qquad (6)$$

From this it can be shown by a tedious though not difficult derivation that the real frequency response is given by :-

$$\left|\frac{V_o}{V_1}\right| = \sqrt{\frac{1 + A^{2(1+N)}B^{2N} - 2A^{(1+N)}B^N\cos wNT}{1 + A^2B^2 - 2AB\cos wT}} \qquad (7)$$

where w = 2πx frequency.

In the special case where A = 1 and B = 1, the relationship simplifies to :-

$$\left|\frac{V_o}{V_1}\right| = \left|\frac{\sin\frac{wNT}{2}}{\sin\frac{wT}{2}}\right| \qquad (8)$$

The obvious difficulty with a recirculating scheme is that caused by unavoidable non-linearities in the record-replay system. Since this system forms a closed loop around which information must circulate, it is unavoidable that any signal will be degraded more and more each time that it traverses the closed loop system. If the overall signal gain around this loop is greater than unity, then the signal will increase in amplitude each time it traverses the closed loop. Eventually the signal becomes so large that saturation in the record-replay system causes severe degradation of the signal.

If, on the other hand, the overall signal gain around the closed loop system is less than unity, then the signal is reduced in amplitude each time that it traverses the loop. Such a feature is actually desirable, since it ensures that more recently recorded information has a larger amplitude than information recorded earlier. However, if a very rapid decay of recorded information is to be avoided, the loop gain must be only very slightly less than unity.

From the mathematical derivation given earlier it can be seen that the input-output response of a sinusoidal recirculating system is zero at some frequencies, so it is not at all constant with frequency. This fact introduces additional difficulties.

7.5 THE EFFECT OF NOISE

It is necessary to consider the effect of noise in a recirculating system of the type discussed above. As seen earlier :-

$$\frac{V_o}{V_1} = A \exp(-sT) \sum_{n=0}^{n=\infty} A^n B^n \exp(-nsT)$$

Now if V_1 is a Gaussian noise waveform, then the series for V_o contains noise components consisting of the noise delayed by nT and multiplied in amplitude by A^nB^n etc. for values of n from 0 to infinity. The R.M.S. value is given by the square root of the sum of the squares of all components. It is of interest to consider the ratio :-

$$\frac{\text{R.M.S. output noise}}{\text{R.M.S. input noise}} = (1 + A^2B^2 + A^4B^4 + \ldots\ldots\ldots)^{\frac{1}{2}}$$

$$= (1 - A^2B^2)^{-\frac{1}{2}}$$

This ratio is plotted against the value of AB in fig. 7.3

It can be seen from this curve that if the loop gain AB approaches unity, then the R.M.S. noise is increased excessively by the positive feedback. Since the loop gain is required to be only very slightly less than unity, this method of using positive feedback in a linear system is not a desirable way of remembering occurrences.

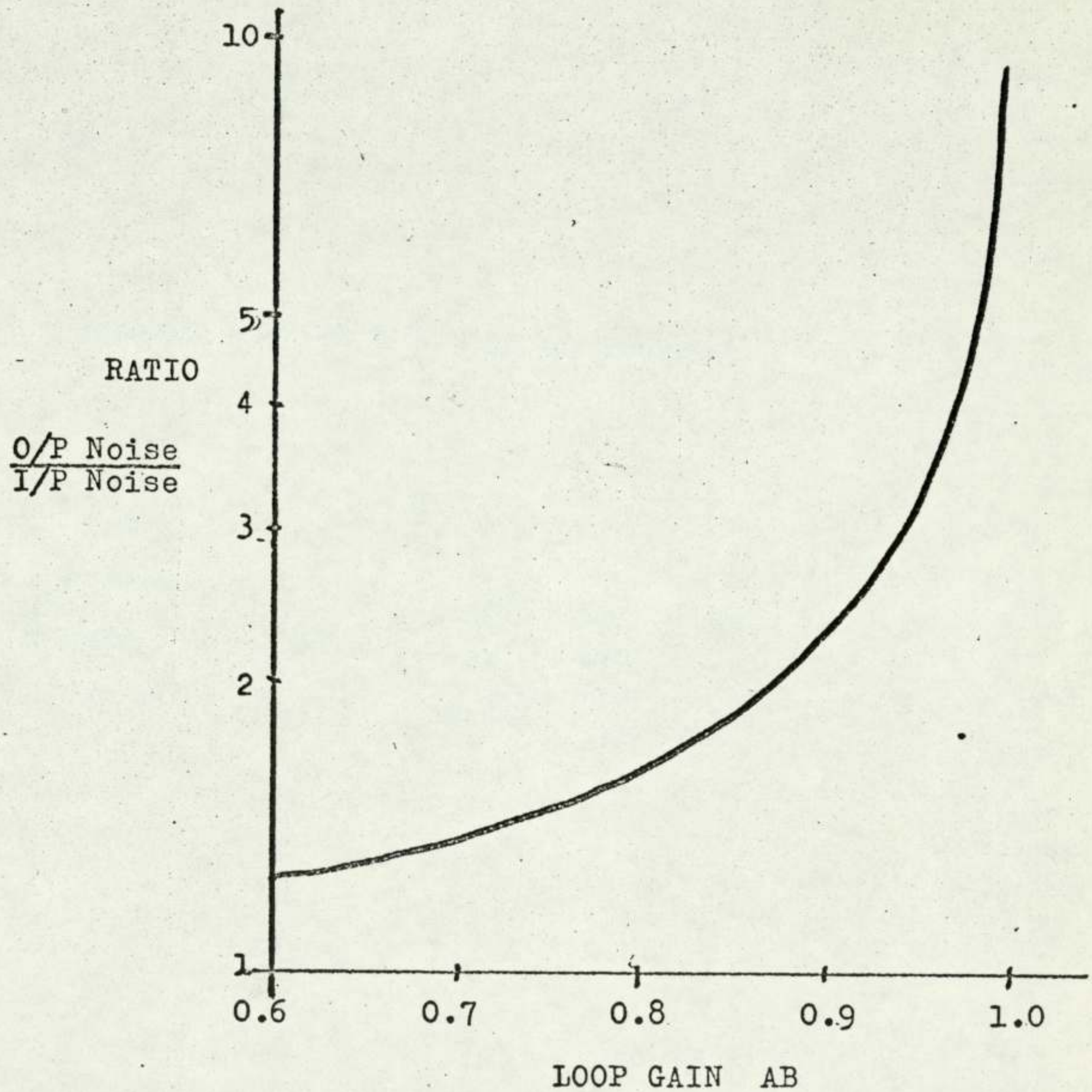


fig. 7.3

VARIATION OF NOISE WITH LOOP GAIN

7.6 PRACTICAL WORK WITH SINUSOIDAL RE-RECORDING

Preliminary investigation of the possibilities of the continuous re-recording method was carried out using two different forms of commercial magnetic recorder. One was a magnetic disc recorder which was intended originally as an early "portable" dictation machine. An additional recording head was fitted to one of these machines, together with additional electronic amplification. This machine had limitations in that the driving motor was clockwork and the amplifiers were battery operated. nevertheless, with this machine preliminary experience of the difficulties encountered with the continuous re-recording method was obtained.

Some further work was carried out using a low-cost commercial magnetic tape deck and amplifiers. With this equipment, which was operated from an A.C. supply, the severe problem which can be caused by interference from external fields was encountered. Even small amounts of interference can be greatly enhanced in a linear recirculating scheme such as this.

The experience gained with these early systems has had a large influence on the form of machines now being investigated by the writer.

7.7 CONCLUSIONS ON CONTINUOUS RE-RECORDING

The experience gained with attempts to use the method of continuous re-recording of linearly added sine waves produced the following conclusions :-

1. It is impracticable simply to recirculate information in sinusoidal form using magnetic recording, since the unavoidable distortion is cumulative. However, the sine waves could possibly be reshaped by the insertion of filters into the loop. This possibility has not yet been fully investigated.

2. Great care is necessary with a recirculating scheme to avoid the cumulative effects of stray induced voltages. Once again, there are possibilities of minimising such effects by use of filters in the loop.

3. It would be necessary to develop some form of fast-acting automatic loop gain control to hold the loop gain very slightly less than unity. There is the possibility here of using a master pilot tone in an automatic gain control system.

4. The selection of frequencies for the primary sources is not easy, and this limits the possible amount of storage in any one system.

8. RECTANGULAR WAVES

8.1 POSSIBLE USE OF CLIPPED SINE WAVES

If rectangular waveforms were used in a recirculating loop information storage system, then it would be permissible to have the gain around most of the loop in excess of unity. Reshaping of the stored waveform to the standard height and shape could then take place at a single point in the storage loop. It should be mentioned that biological storage and transmission systems make use of continuous reshaping of information during transmission, rather than reshaping at a single point.[23]

In a system including reshaping at a single point in the recirculating loop, it would not matter at all if the storage loop had such a frequency response that the harmonic content of the stored rectangular wave was almost completely removed, leaving only the fundamental sine wave. Such a sine wave could then be "clipped" in order to reproduce the original rectangular wave. This clipping could be done by the use of various non-linear circuit elements, for example by use of a Zener diode.

If a sine wave is clipped as shown in fig. 8.1 then if the sine wave is $v = V \sin wt$, of peak amplitude V, and it is clipped at an amplitude of $V \cos\theta$, then the angle of clipping must be $\frac{\pi}{2} \pm \theta$

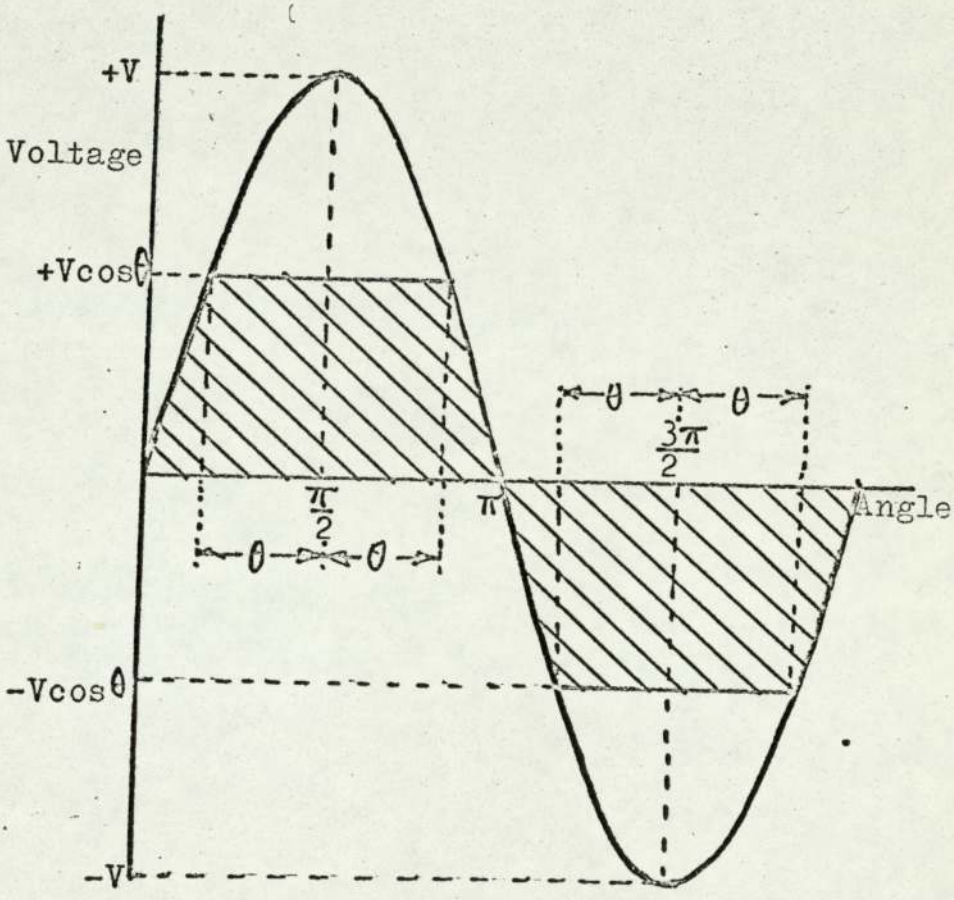


fig. 8.1

CLIPPED SINE WAVE

Expansion of such a waveform gives :-

$$v = \frac{2V}{\pi}\Big[\left(\frac{\pi}{2} - \theta + \frac{1}{2}\sin 2\theta\right)\sin wt$$
$$+ \left(\frac{1}{2}\sin 2\theta - \frac{1}{4}\sin 4\theta\right)\frac{1}{3}\sin 3wt$$
$$- \left(\frac{1}{4}\sin 4\theta - \frac{1}{6}\sin 6\theta\right)\frac{1}{5}\sin 5wt$$
$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\Big]$$

The factors $\frac{1}{x}\sin x\theta$ appear here. The approach to a true rectangular wave is determined by the degree of invariance with frequency of the values of the differences between these factors. The values of the factors are plotted against the value of x for values of θ of 85° and 89° in fig.8.2 It is seen that the values approach constancy as the clipping level is decreased.

At least one of the harmonics contained in the clipped voltage waveform has a zero amplitude.[24] This fact could be used for sensitive control of the clipping level. It is of interest to note that as the clipping level is decreased, so the order of harmonic frequency at which zero amplitude is achieved is increased. The way in which this order increases as the clipping level is decreased is shown in fig.8.3

This method of clipping and filtering is satisfactory for use as a non-frequency sensitive control of amplitude of a single sinusoidal frequency.

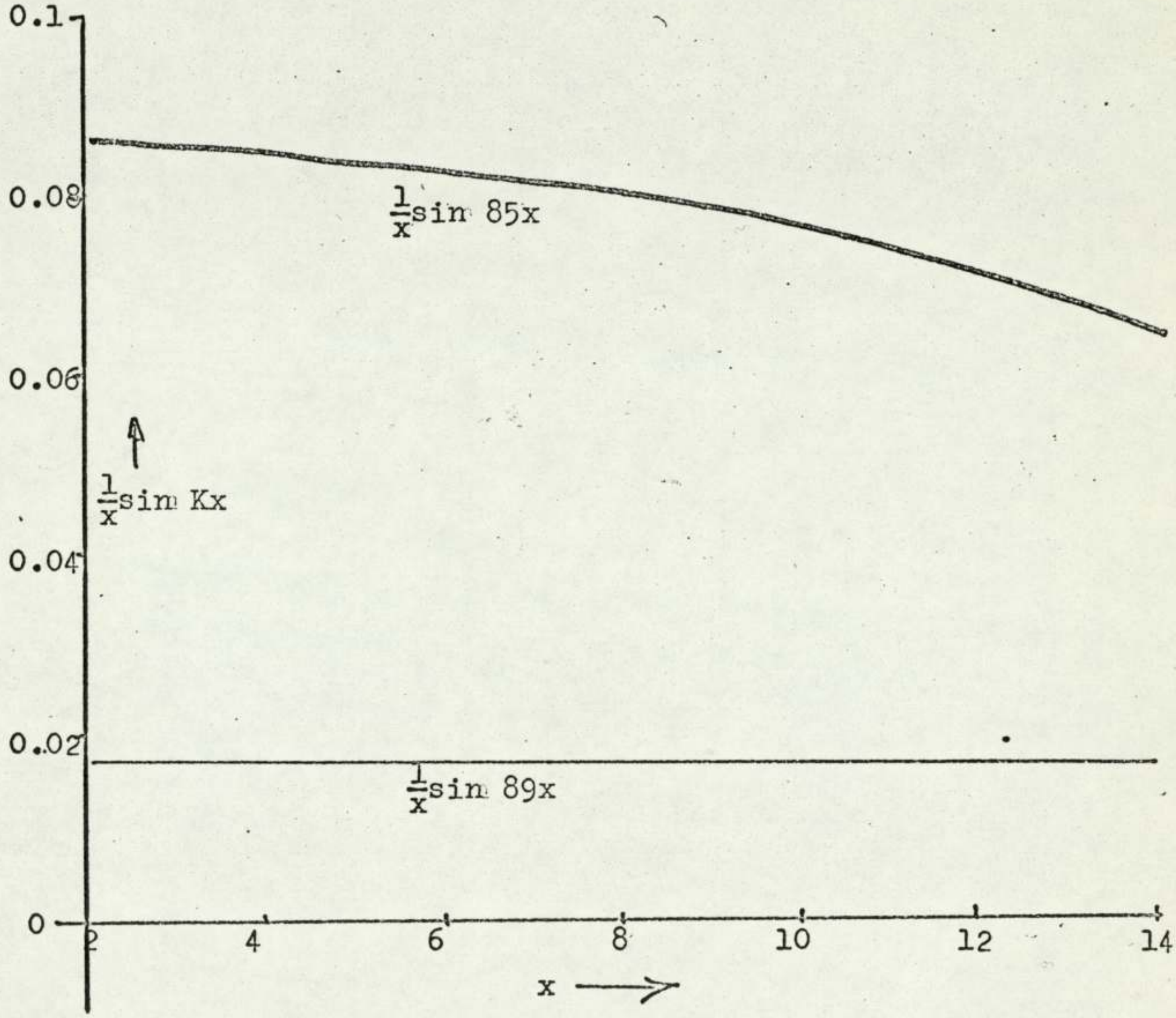


fig. 8.2

VARIATION OF FACTORS OF EXPANSION

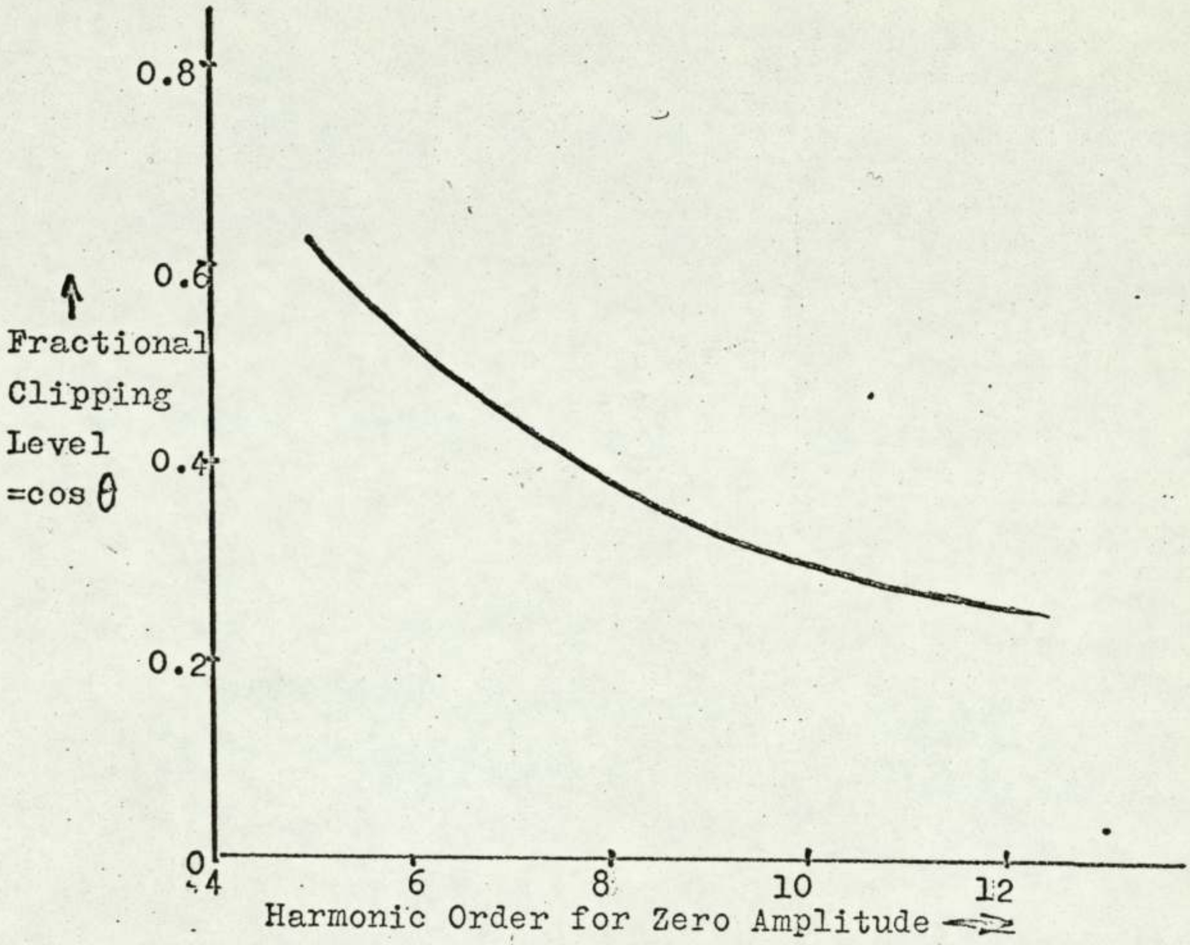


fig. 8.3

HARMONIC ORDER FOR ZERO AMPLITUDE

Indeed, at very low signal frequencies there is no really satisfactory alternative method of amplitude control. However in general it is not suitable for use with complex waves which effectively consist of the summation of a number of sine waves. The fundamental reason for this is that amplitude limitation of a general complex wave changes the distribution and the amplitudes of the constituent sine waves.

Such problems might not exist if rectangular waves were used exclusively throughout the system, and the loss of harmonic content was minimised. Reshaping of a wave containing a number of rectangular waves, while retaining all of the information contained in the waves, would appear at first sight to be a much more manageable task than is the reshaping of more complex waves containing a general distribution of sine waves. It is therefore of interest to consider some of the properties of mixtures of rectangular waves.

8.2 LINEAR ADDITION OF RECTANGULAR WAVES

If two rectangular waves having equal amplitudes but different repetition frequencies are added together linearly, then the resultant contains three levels as shown for example in fig.8.4 Since the linear sum is being taken, the absolute level of the waveform is unimportant. However, this level has an effect if there is any non-linearity in the addition process.

A further point of interest is that if the repetition periods are prime numbers, then the repetition period of the sum waveform is equal to the product of the repetition periods of the two constituent rectangular waveforms. Thus in the example, the repetition period of the first waveform is 5 units, the repetition period of the second waveform is 3 units, and the repetition period of the sum waveform is 5 x 3 = 15 units.

The linear sum of n rectangular waves of equal amplitudes but differing frequencies contains (n+1) levels of amplitude. In fig.8.4 the value of n is only equal to 2. However, the waveform of a linear sum wave rapidly becomes more and more complex as the number of components increases. Such a wave would not be suitable for re-shaping by use of simple amplitude clipping arrangements, and so this method of linear addition is not suitable on this count.

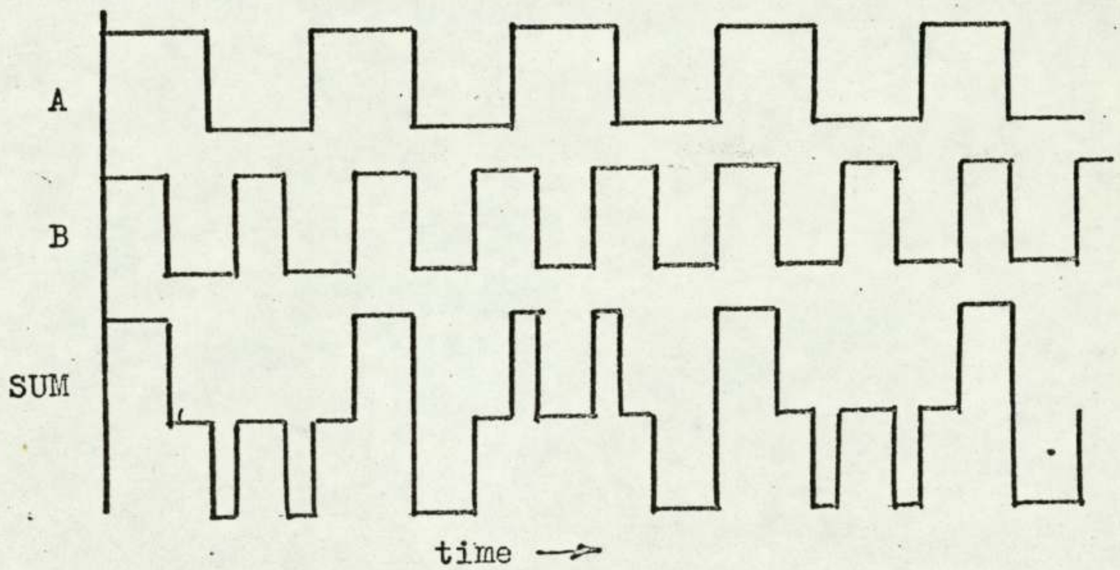


fig. 8.4

LINEAR ADDITION OF RECTANGULAR WAVES

The outstanding reason why simple linear addition of rectangular wave signals cannot be considered is that the maximum overall amplitude which can occur is equal to n units if a number n of rectangular waveforms, each of unit amplitude, is added together linearly. This linear addition would lead to saturation in any practical recording medium. Such an effect would cause recorded information to be lost without hope of recovery. Thus although reshaping of summated rectangular waves is not impossible, this will not prevent loss of information if simple linear addition is used.

8.3 LOGICAL OPERATION ON RECTANGULAR WAVES

It has been seen that the simple arithmetic summation of a number of rectangular waves produces a resultant output wave which has a number of different amplitude levels and which is therefore not very suitable for re-shaping by use of non-linearities. It is consequently of interest to consider as an alternative the combination of rectangular waveforms by use of methods which are inherently incapable of producing more than two discrete output signal amplitude levels. One possible method is by generation of the logical product of rectangular waves, for example by use of logical gate circuits.

Consider two unidirectional rectangular waves v_1 and v_2, having identical amplitudes V but differing angular repetition frequencies w_1 and w_2. Now if these two waves are applied to the inputs of a two-way logical AND gate, then the output of the gate is given by the following table :-

v_1	v_2	O/P
0	0	0
0	V	0
V	0	0
V	V	V

This table should be compared with a table of instantaneous products of v_1 and v_2. The products can each have one of only two different values, 0 and V^2.

v_1	v_2	Product
0	0	0
0	V	0
V	0	0
V	V	V^2

From this it is seen that the logical AND gate gives the required output, modified only by a scale factor.

If v_1 and v_2 are rectangular waves then the product v_1v_2 is given by the expansion :-

$$\begin{aligned} v_1v_2 = {} & V^2/4 \\ & + V^2/\pi\Big[\sin w_1t+(1/3)\sin 3w_1t+(1/5)\sin 5w_1t+\ldots. \\ & \quad +\sin w_2t+(1/3)\sin 3w_2t+(1/5)\sin 5w_2t+\ldots.\Big] \\ & +4V^2/\pi^2\Big[\sin w_1t.\sin w_2t +(1/3)\sin w_1t.\sin 3w_2t \\ & \quad +(1/5)\sin w_1t.\sin 5w_2t +\ldots\ldots. \\ & \quad +(1/3)\sin 3w_1t.\sin w_2t +(1/9)\sin 3w_1t.\sin 3w_2t \\ & \quad +(1/15)\sin 3w_1t.\sin 5w_2t +\ldots\ldots.. \\ & \quad +(1/5)\sin 5w_1t.\sin w_2t +(1/15)\sin 5w_1t.\sin 3w_2t \\ & \quad +(1/25)\sin 5w_1t.\sin 5w_2t +\ldots.. +\ldots..\Big] \end{aligned}$$

Here the first term is a constant. The second term is simply formed from the algebraic sum of the two original rectangular waves.

If the steady component was removed from the original rectangular waves so that :-

$$v_1 = \frac{2V}{\pi}\left(\sin w_1t + (1/3)\sin 3w_1t + (1/5)\sin 5w_1t +\ldots\right)$$

$$v_2 = \frac{2V}{\pi}(\sin w_2t + (1/3)\sin 3w_2t + (1/5)\sin 5w_2t + \ldots$$

then the product wave would only contain the third set of terms in the expansion above, the first term and the second set being eliminated entirely.

Unfortunately, removal of the steady component of the original waves changes the possible instantaneous levels from V and 0 to $+\frac{V}{2}$ and $-\frac{V}{2}$. Consequently the table of instantaneous products becomes :-

v_1	v_2	Product
$+V/2$	$+V/2$	$+V^2/4$
$+V/2$	$-V/2$	$-V^2/4$
$-V/2$	$+V/2$	$-V^2/4$
$-V/2$	$-V/2$	$+V^2/4$

The action corresponding to this table can be obtained by use of an inverted exclusive-OR arrangement as shown for example in fig. 8.5 though this is not the only way to implement such a "logical product" table.[25]

The writer has used such an arrangement as a generator of triangularly pulse-width-modulated test signals for p.w.m. amplifier systems.[26] A convenient graphical way of visualising the effects of input frequency changes was shown to be the use of a "chess-board" representation of the two inputs. In this the two axes represent the rectangular wave inputs and shaded areas represent regions where both inputs are simultaneously positive. Various effects can then be shown.

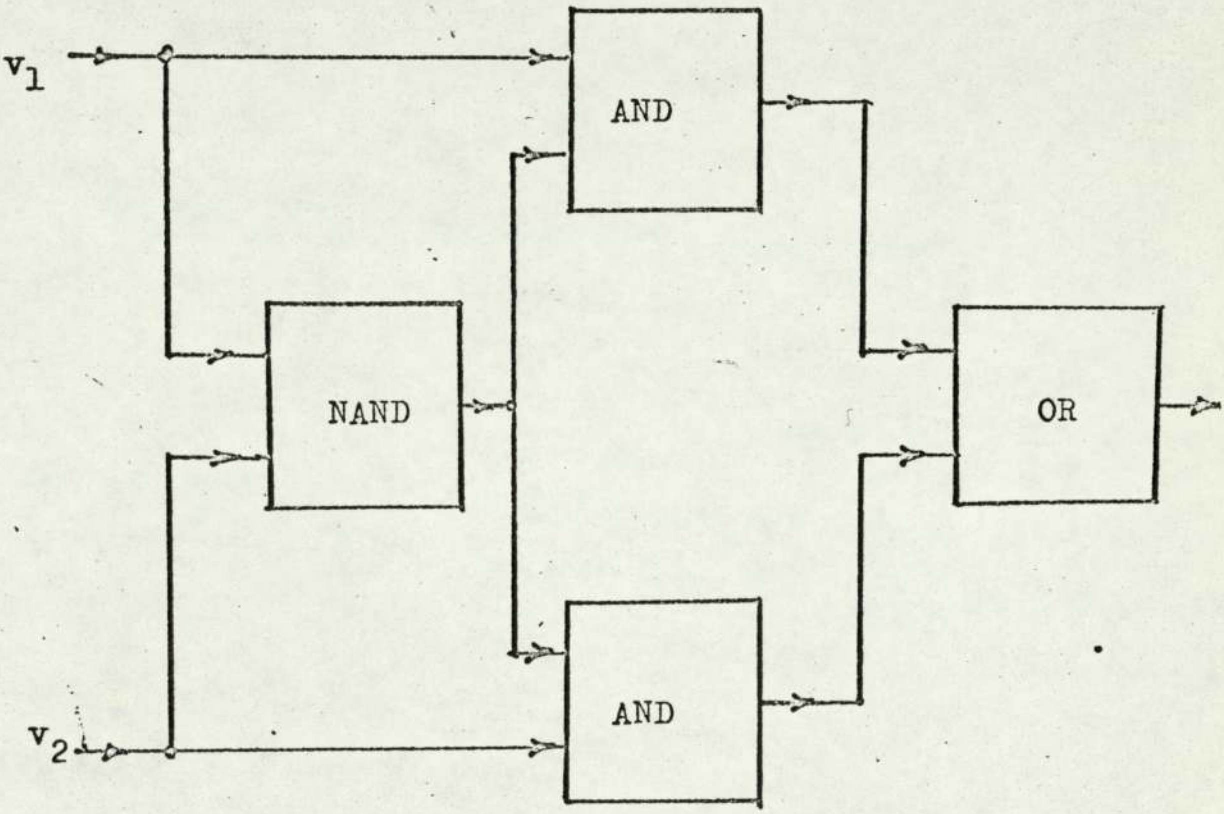


fig. 8.5

EXCLUSIVE-OR CIRCUIT

8.4 USE OF A GATE AS A SELECTIVE FILTER

If rectangular waves are mixed to obtain the logical product as suggested earlier, then this logical product waveform contains the required past-association information for application to the memory. If this information is stored, then provision must be made for its utilisation at a future time.

If the occurrences of inputs A and B have coincided in the past, then excitation of the rectangular wave generators having frequencies w_A and w_B has coincided in the past. These rectangular waves have been mixed and the results have been stored. The stored information takes the form of a triangularly pulse-width-modulated wave.[26] Now the problem must be considered of how this stored information is to be used in the future.

If one of the inputs, say A, occurs alone in the future, then it must be possible to make use of this input A, together with the information stored in the memory about the past association of input A with input B, to produce excitation of the output b which corresponds to input B. In this way the remembered past association of A and B can be used to cause excitation of a particular output b, not only by the occurrence of its corresponding input B but also by the occurrence of the associated input A.

It would be possible to use tuned circuits as selective output filters, together with some form of mixing device as envisaged in the earlier sinusoidal scheme. However, there would be several disadvantages in the use of such a method. The schemes discussed here are envisaged for eventual use with a very large number of inputs and outputs. The resulting large bank of filters would be bulky, since inductances would be required. The inductances would have to be shielded in order to prevent effects caused by extraneous fields. If a magnetic recorder was used as the memory device, the speed would have to be closely controlled. Any change of inductance due to drift would introduce additional problems if a very large bank of filters was involved.

As an alternative to the use of tuned filters, logical selective gates can be used. As an example, a two-input AND gate with one input connected to a master frequency pulse supply will only give a large mean value of output if the other input is connected to a source producing pulses at the same frequency. Any difference between the two frequencies will cause beats to appear and there will be a reduction of the mean output from the gate. Thus such a gate can be used as a selective filter if one input is connected to a master pulse source.

8.5 RECTANGULAR WAVE ANALYSIS

Consider a recorded triangularly pulse-width-modulated wave as produced by a simple AND gate to be applied to one input terminal of a further logical AND gate. The other input terminal is connected to a test source producing a rectangular wave at a repetition frequency almost equal to that of one of the original constituent rectangular waves. Then the mean output of the AND gate varies slowly at the difference frequency between the recorded frequency and the newly-applied frequency.

The frequency of the rectangular wave from the test source can be made exactly equal to the frequency of one of the originally associated rectangular waves contained in the recorded triangularly pulse width modulated wave. The output of the gate is then reduced to zero provided that the test rectangular wave is exactly in antiphase to, as well as equal in frequency to, one of the original rectangular waves. An example of this action is shown in fig.8.6 Here the original associated signals had repetition periods of five units and of three units respectively. The logical addition of these two waveforms produces the width modulated waveshape shown. The logical addition of this logical sum waveform to the negative of the rectangular wave having a period of three units then gives a zero output

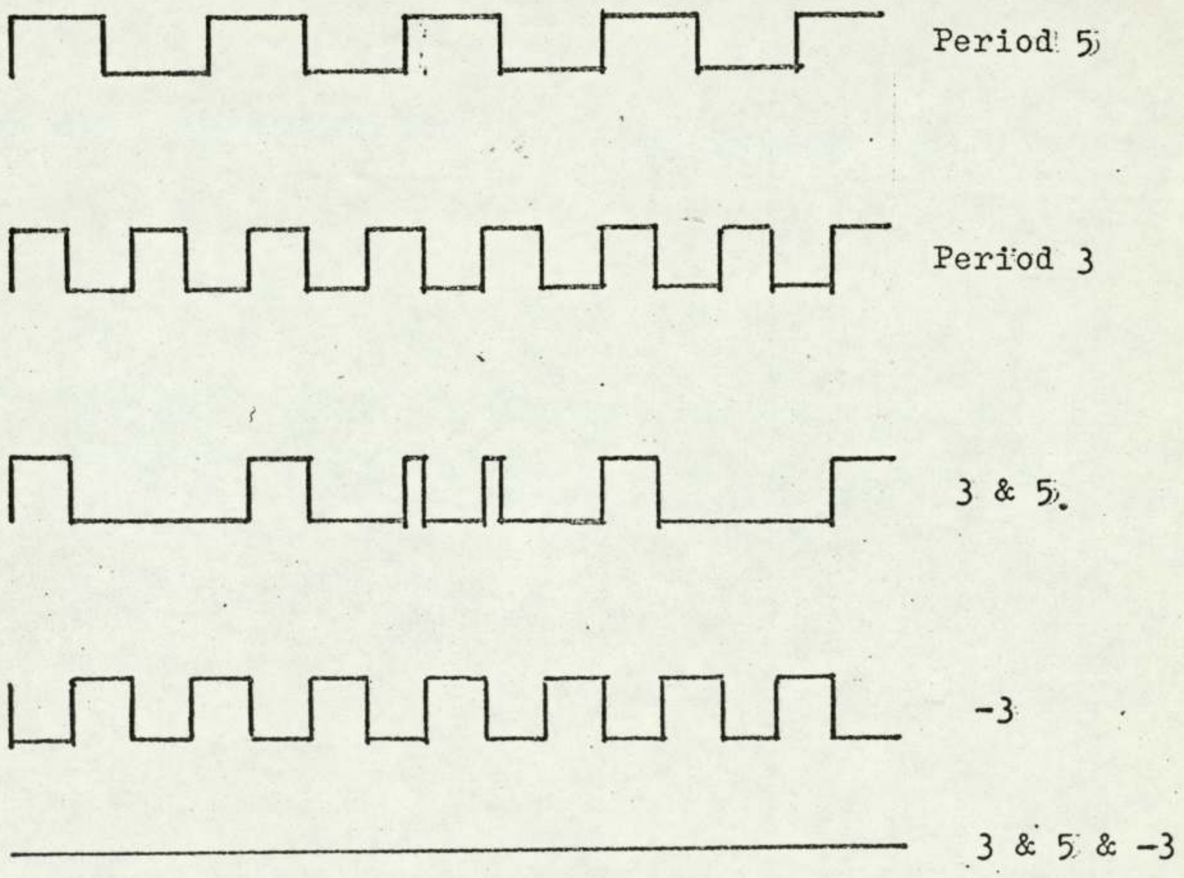


fig. 8.6

RECTANGULAR-WAVE ANALYSIS

from the AND gate as shown. Similarly, the application of the negative of the rectangular wave having a period of five units will give a zero output. On the other hand, the application of a rectangular wave with a period of for example four units gives an output as shown in fig. 8.7.

Detection of the null can be used in a very sensitive selective filter arrangement. In the present case, when a particular frequency is applied, such an arrangement can be used to detect the presence of a recorded association with that frequency. In logical terms, the process can be expressed as :-

$$(A.B.C)\overline{B} = B.\overline{B} = 0$$

However, it is not sufficient simply to determine that there has been an unspecified association in the past. It must be possible to determine the actual frequency of the rectangular wave which has in the past been associated with the presently impressed wave.

Now a triangularly pulse-width-modulated wave can be formed by the logical addition of two rectangular waves neither of which has any steady content. It is convenient here to consider a simple specific example. In fig.8.8 two waves having repetition periods of five units and three units respectively, and each having a mean value of zero, are logically added in an "exclusive-OR" circuit to produce the wave shown. This wave is an elementary form of pulse-width-modulated wave. Now

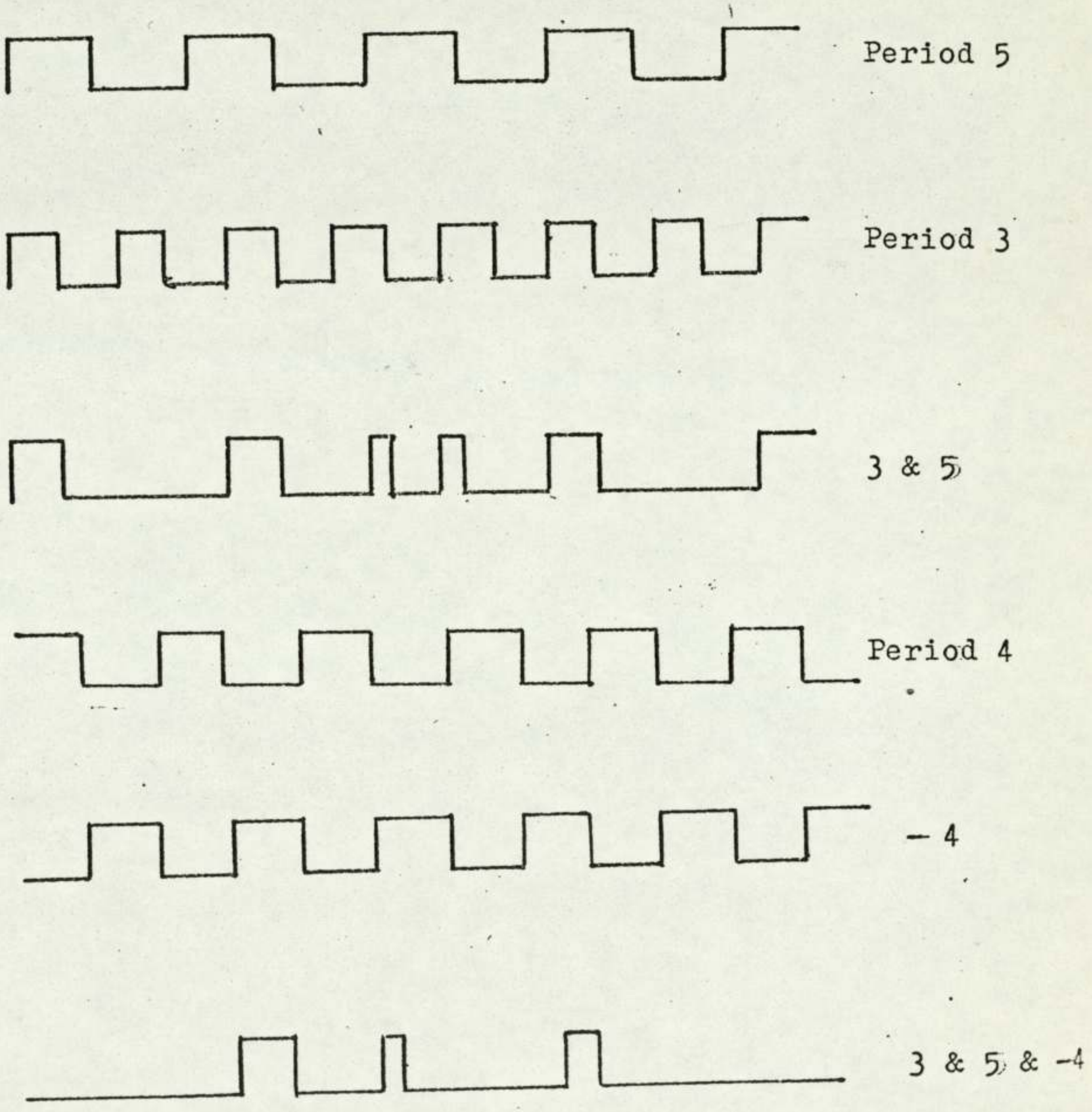


fig. 8.7

EFFECT OF APPLICATION OF WRONG FREQUENCY

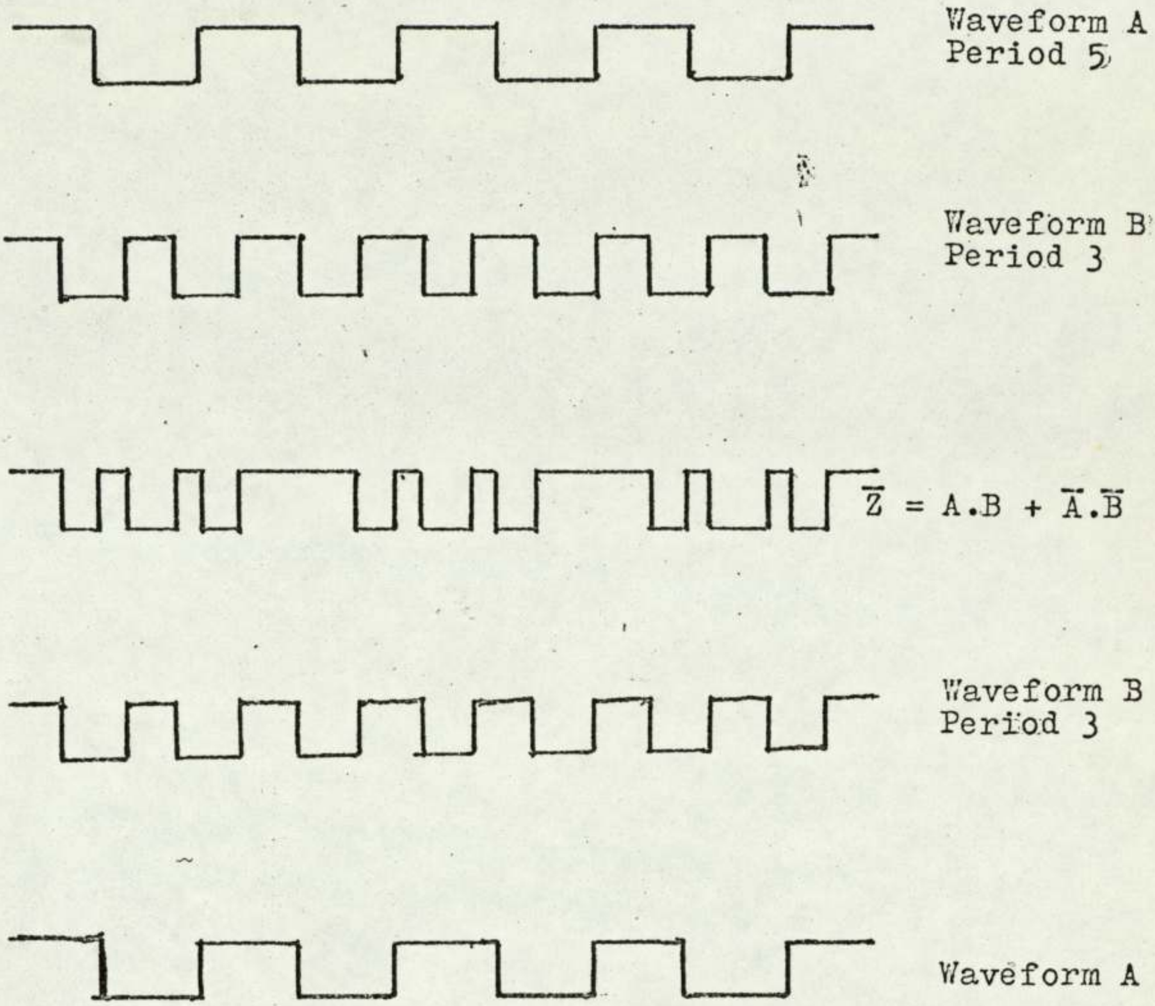


fig 8.8

USE OF EXCLUSIVE-OR CIRCUIT

suppose that this wave is applied to one terminal of a further exclusive-OR circuit, while a zero-mean-content rectangular wave of repetition period three units is applied to the other input terminal. The output of the exclusive-OR circuit is then the associated rectangular wave of five units period as shown in fig. 8.8. The logic statement of this approach is :-

let $\quad A.\overline{B} + \overline{A}.B = Z$

then $\quad B.\overline{Z} + \overline{B}.Z = B(A.B + \overline{A}.\overline{B}) + \overline{B}(A.\overline{B} + \overline{A}.B)$

$$= A.B + A.\overline{B}$$

$$= A \qquad \text{as required.}$$

It should be noticed, however, that the exclusion of unwanted output is not at all complete with this method. While the correct associated waveshape is always obtained when an associated signal is applied, non-associated signals are not prevented from producing an unwanted output. This output will not be a periodic rectangular wave such as those given by the correct associated signals, and it can therefore be eliminated by further filtering.

It is of interest to compare the mixing of rectangular waves as discussed here with the optical work which has been carried out on Moiré fringes and their applications.[27]

8.6 DEMODULATION WITH SYMMETRICAL RECTIFIER

In the course of work on the methods discussed here, it is necessary to investigate the form of the recorded width-modulated pulses. In the methods considered, the information about past associations is stored on the magnetic recorder in the form of pulses of variable widths. As has been seen, the association of two rectangular waves results in the production of a triangularly pulse-width-modulated wave. Demodulation of such waves is commonly achieved by the use of simple low-pass filters. While this is adequate for many audio applications, there is a severe loss of information at the higher frequencies, and for example a triangular wave is degraded in such a filter to become a waveform little different from a sine wave. It was felt that better methods should be available in the present research. Accordingly, the possibilities of the method of symmetrical rectification were investigated.

The writer has shown elsewhere[28] that a device having a perfectly symmetrical though non-linear characteristic such as that of fig.8.9 is suitable for use as a rectifier of pulse waveforms. If a waveform such as that of fig.8.10 with an on-time of t_a and a repetition period of t_k and having no D.C. content, is applied to an ideal rectifier of this type, the mean value of the output of the fectifier is

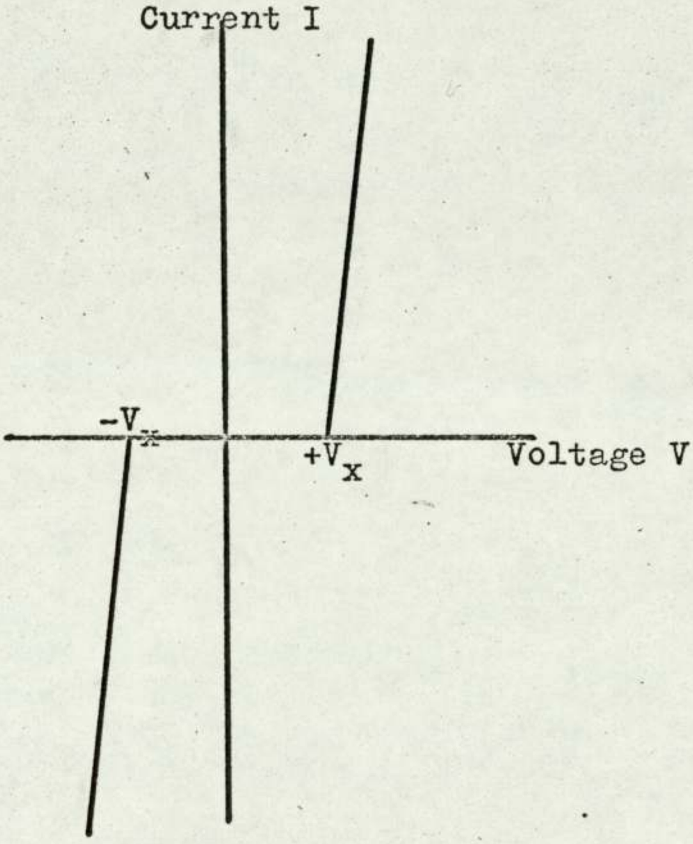


fig. 8.9

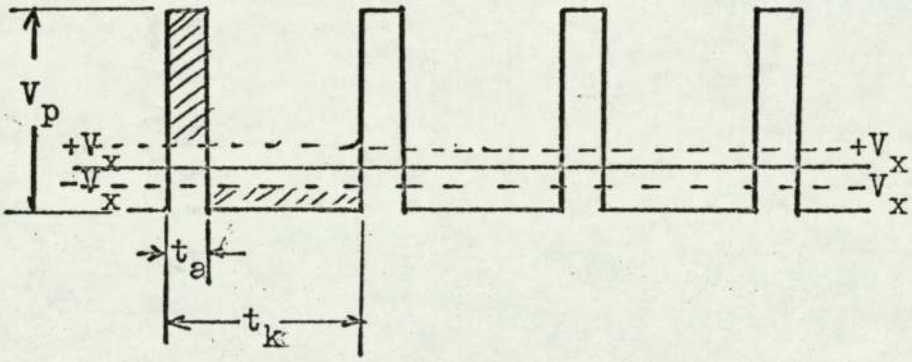


fig. 8.10

given by the linear relationship :-

$$V_{mean} = V_x \left(1 - 2\frac{t_a}{t_k}\right)$$

It is assumed here that both the maximum positive and the maximum negative levels of the waveform exceed the clipping level V_x of the symmetrical rectifier. This assumption is fulfilled provided that the peak-to-peak height V_p of the rectangular waveform is greater than :-

$$V_p > \frac{V_x t_k}{t_a}$$

Within this limitation, the mean output of the symmetrical rectifier is proportional to the mark-to-period ratio of the applied rectangular wave. Thus the relationship is linear, passing through zero at the point where $\frac{t_a}{t_k} = 0.5$, as shown in fig. 8.11.

Thus the symmetrical rectifier is an ideal device for use as a demodulator of pulse-width-modulated signals.

In the present investigation, the symmetrical rectifier is of use for the conversion of the triangularly pulse-width-modulated pulses into a triangular waveform. Such a converting device is required in the equipment considered here. The necessity for the incorporation of such a conversion is discussed below.

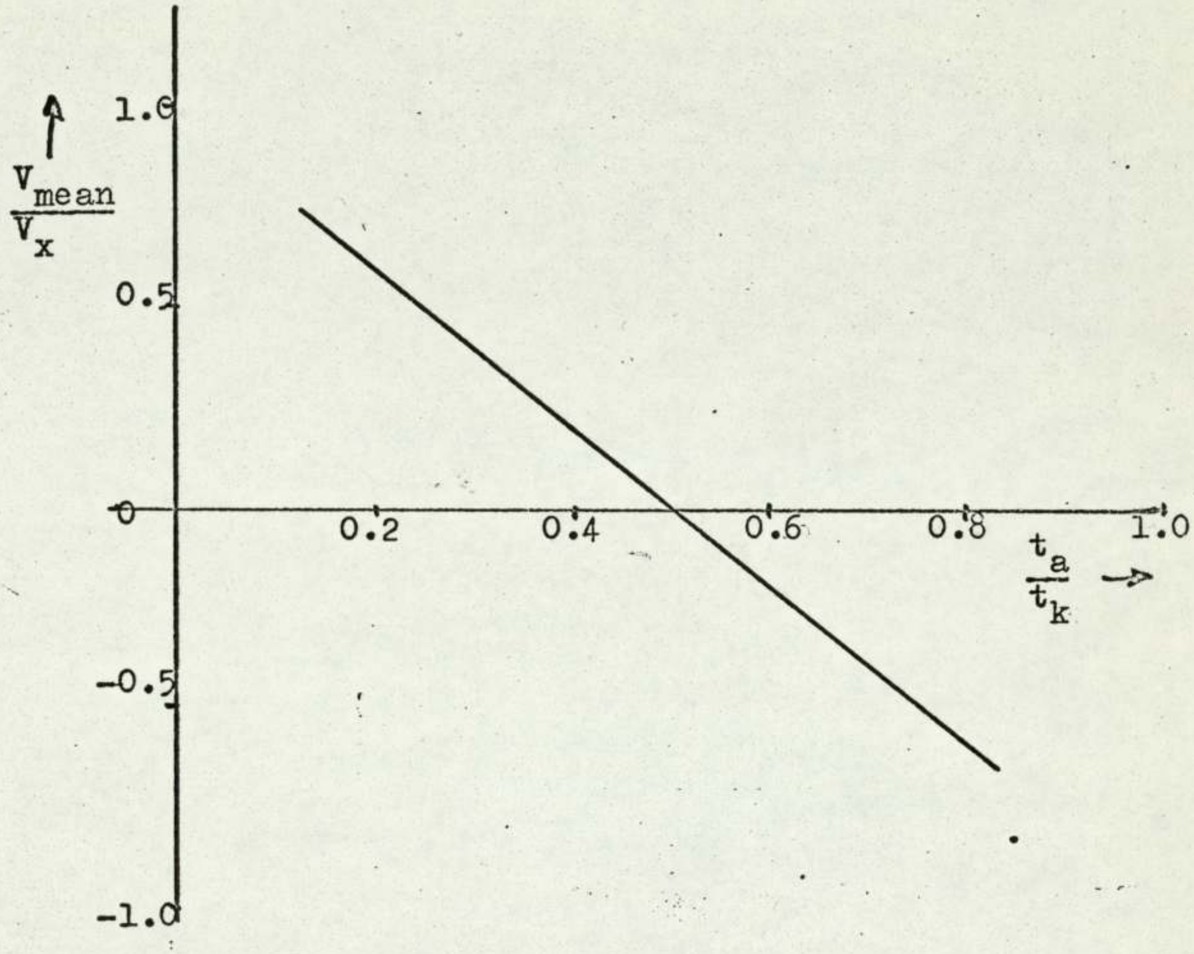


fig. 8.11

CHARACTERISTICS OF SYMMETRICAL RECTIFIER

8.7 BASIC DIFFICULTIES WITH PULSE-WIDTH-MODULATION

If pulse-width-modulation is to be used in the memory of an associating type of learning machine, there are certain basic practical difficulties which must be overcome. It is possible to regenerate and to recirculate the stored information, though some degradation must be expected due to shifts of the edges of the pulses during the regeneration process. Fortunately, both leading and trailing edges of the pulses should be shifted by the same amount in a purely linear system. However, this will not be true if the system has change-over characteristics which depend on the direction of change.

Such a ddpendence could be caused for example by the cutting-off of an emitter-follower, using an NPN transistor, which occurs because of stray capacitance across the load when a negative-going edge appears. More fundamentally, carrier storage in semiconductors is likely to introduce such undesirable asymmetrical effects. Even small shifts are likely to be exaggerated in any scheme featuring dynamic recirculation of information in pulse form. It is noteworthy that biological systems which feature continuous pulse re-shaping do not seem to suffer from such defects.

Recirculation degradation is an unfortunate effect. However, an even more important problem is caused by the difficulty of adding new information to a variable pulse-width recording. In order to do this, it would be necessary to modify the width of the pulses to give the algebraic sum of the width of the recorded pulses plus the width of the incoming pulses plus a constant. One possibility here is the use of the symmetrical rectifier to produce triangular waveforms from the pulse-width-modulated recorded information, possibly followed by filters, before the point at which the new information is added-in.

8.8 CONCLUSIONS ON THE USE OF PULSE-WIDTH-MODULATION

The work described here on the use of pulse-width-modulated recording in a cybernetic association-memory scheme led to the following conclusions :-

1. A great deal of further work is required to achieve successful memory devices incorporating complex pulse-width-modulated waves.

2. Normal pulse generators do not produce a sufficiently stable frequency and, more important, the tape speed of a recorder is not sufficiently stable to ensure a close matching of speed and frequency. It is preferable to derive all pulse signals from a master track recorded on the tape, so that speed and frequency will vary together. Such a scheme is being incorporated in the further work on association memory devices now being carried out.

3. If tape loops are used in association-memories, special arrangements are required to overcome the loss of signal at the loop splice in the tape. Possibly magnetic disc or drum recording could reduce this difficulty, though it would still be necessary to have a reference "start" point in the recording. Such arrangements in various forms are incorporated in the further work now in progress.

It was realised that it is not necessary to record the actual pulse-width-modulated waves on the magnetic medium.. The change-over times of the pulses convey all of the information contained in a wave, since the amplitude of the wave is constant. In fact, if a pulse wave is applied direct to a magnetic recording head, the reproduce head merely gives a short differentiated pulse each time that the recorded pulse changes polarity. A bistable circuit can reconvert these short pulses to pulse-width-modulated form.

9. POSSIBLE USE OF SHORT PULSES

9. POSSIBLE USE OF SHORT PULSES

So far, the possibilities of using sinusoidal signals and of using rectangular wave signals in recirculating cybernetic memories have been discussed. In the latter case, while the problems introduced by distortion effects in the earlier system can be overcome, difficulties are inherent in the process for the addition of new association information to the earlier-remembered information. In addition, the application of a rectangular wave of current to a magnetic recording head produces only short pulses corresponding to the changeover points at the replay head, since a steady flux can give no signal.

These two facts suggest the possibility of the use of short pulses in a cybernetic memory storage system. The similarity of such an approach to the use of short pulses which occurs in biological nervous systems is noteworthy. It is essential in cybernetic work to take care not to attempt the excessive use of such analogies. However, it is of interest to consider some known aspects of biological systems and the consequent feasibility of adaptation of some of these basic features into engineering devices.

10. NEURAL NETWORKS

10.1 NEURAL PROPAGATION

At this point some of the known features of neural impulse propagation should be summarised. A nerve fibre passes signals on an all-or-nothing basis. The neural sheath is normally non-conducting. When the sheath does conduct, it does so for a short time, of the order of a millisecond, only and then it recovers to the non-conducting state. Any in-between condition of partial conduction can persist for a very short time only. Information is conveyed along nerve fibres by such transients in a form of pulse-frequency-modulation.

The nerve fibre has very special properties which make its action unlike that of an electrical passive transmission line. The form of the propagating nervous signal depends only on the nature of the nerve fibre and not on the means by which the signal is started. After an impulse has passed a given point on the nerve fibre, a short recovery (or refractory) time is required before the next impulse can pass. If separate discharges start from opposite ends of a fibre, they are mutually destructive when they meet or collide, since each leaves behind it a refractory region through which neither impulse can continue to propagate. There can be no reflection of a signal when it reaches the end of a neural transmission line, because of the refractory region behind it.

Biological nerve fibres convey information by means of pulse-frequency-modulation. The greater the excitation intensity E, the greater the frequency of the pulses produced. Let I be the (constant valued) intensity of a single impulse and let f be the rate or frequency of the impulses. Then we can write :-

$$E = If \qquad (1)$$

Now a neuron produces impulses when it is stimulated by some external stimulus S, provided that the stimulus S exceeds a certain threshold intensity i. As a first approximation we can write :-

$$f = K_1(S - i) \qquad (2)$$

where K_1 is a constant. Hence :-

$$E = IK_1(S - i) \qquad (3)$$

$$E = K(S - i) \qquad (4)$$

where K is a constant, since I is a constant.

However we know from biological observation that there is a maximum possible value of f, so that we can write $f \leqslant F$. A suitable approximation[22] to the value of f is therefore given by :-

$$f = F\left\{1 - \exp\left(- \frac{K(S - i)}{F}\right)\right\} \qquad (5)$$

Hence $$E = IF\left\{1 - \exp\left(- \frac{K(S - i)}{F}\right)\right\} \qquad (6)$$

Now assume that at the end of a fibre, the stimulus conveyed to another neuron is given by :-

$$S_2 = K_2E_1 - T_k \frac{dS_2}{dt} \qquad (7)$$

where K_2 and T_k are constants. This gives :-

$$S_2 = K_2E_1\left\{1 - \exp\left(-\frac{(t - t_p)}{T_k}\right)\right\} \qquad (8)$$

on the assumptions that E_1 is constant and that $S_2 = 0$ when $t = t_p$. The excitation takes a time t_p to travel along the fibre. The first fibre excites the second neuron when $S_2 \geqslant i_2$. This occurs when :-

$$\exp\left(\frac{t - t_p}{T_k}\right) = \frac{K_2E_1}{K_2E_1 - i_2} \qquad (9)$$

Hence it follows that :-

$$t = t_p + T_k \ln K_2K(S - i_1) - T_k \ln (K_2K(S - i_1) - i_2) \qquad (10)$$

This expression gives the time at which the second neuron is excited. It should be noted that the delay is a function of S.

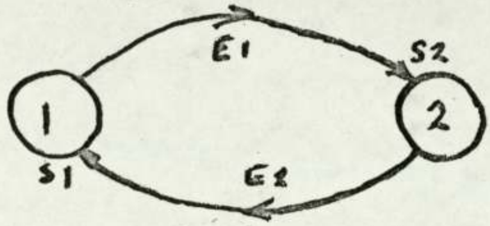
10.2 SIMULATION OF NEURAL PROPAGATION

Many attempts to simulate neural transmission by use of electrical analogues have been made. Lillie[29] used a model made of iron wire immersed in nitric acid. Walter[8], who himself has constructed analogues using neon tubes, criticised the Lillie model on the grounds that the nature of the insulating film in the iron-wire model is almost as mysterious as is the nature of natural nerve fibre. Walter used a cross-coupled arrangement of neon tubes which could be used to demonstrate the neural type of propagation visually. Burns[30] used thyratron valves in a model neuron. Freygang[31] suggested the use of a resistance-capacitance transmission line with voltage pulse regenerators situated periodically along the line. Crane[32] considered various practical forms which artificial nerve fibres can take. Nagumo et al.[23] and Yoshizawa et al.[33] have used the properties of tunnel diodes to provide the necessary pulse regeneration in artificial nerve fibres, while Cote[34] and Rosengreen[35] have used four-layer semiconductor devices.

There has thus been much work on the simulation of nervous transmission of information. There has not been the same emphasis on the use of closed neural loops for temporary information storage, though it seems that such a mechanism may possibly be used in biological systems.

10.3 CLOSED NEURAL LOOPS

It has often been suggested that some memory storage in neural systems uses the properties of closed neural loops.[37] While it is most unlikely that permanent memory takes this form, judging from the non-permanent results of such occurrences as concussion, it is quite possible that temporary information storage takes place in recirculating neural loops. A typical loop is illustrated below.



A neuron having an excitation E_1 excites another neuron with excitation S_2 . On the previously adopted assumption, we can write :-

$$S_2 = K_2E_1 - T_k \frac{dS_2}{dt} \qquad (11)$$

Now if a further fibre carries excitation from the second neuron back to the first, then we can write :-

$$S_1 = K_1E_2 - T_k \frac{dS_1}{dt} \qquad (12)$$

Substituting the earlier approximate forms for E_1 and E_2 we obtain :-

$$S_2 = K_2IF\left\{1 - \exp\left(-\frac{K(S_1-i_1)}{F}\right)\right\} - T_k\frac{dS_2}{dt} \qquad (13)$$

and :-

$$S_1 = K_1 IF\left\{1 - \exp\left(-\frac{K(S_2-i_2)}{R}\right)\right\} - T_k \frac{dS_1}{dt} \quad (14)$$

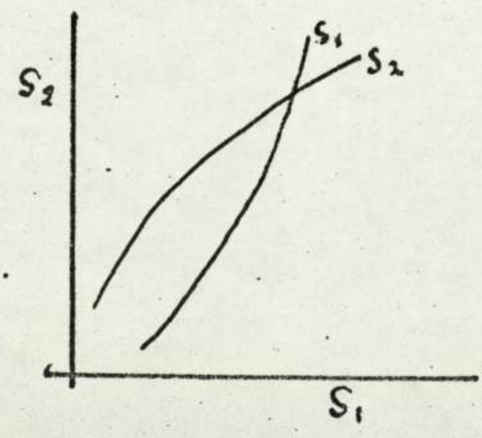
Now it is of interest to consider the conditions for unchanging S_2 and for unchanging S_1. The first is :-

$$S_2 = K_2 IF\left\{1 - \exp\left(-\frac{K(S_1-i_1)}{R}\right)\right\} \quad (15)$$

This equation defines a curve on the S_2, S_1 plane. A similar curve is defined by the other condition :-

$$S_1 = K_1 IF\left\{1 - \exp\left(-\frac{K(S_2-i_2)}{R}\right)\right\} \quad (16)$$

If the two curves in the S_2, S_1 plane intersect, then a stable operating condition exists in which an impulse can continue to circulate around the neural loop once it is started. It is important to note that the actual exponential forms assumed are not essential for the results obtained to be true. It is merely essential that the actual algebraic forms make intersection possible, so that a stable solution, involving the storage of information, exists. The situation is illustrated below.



11. RECIRCULATION

11.1 RECIRCULATING PULSES

As shown in section 7.4 a general expression for the transfer function of a recirculating positive feedback system is given by :-

$$\frac{V_o}{V_1} = A \exp(-sT) \sum_{n=0}^{n=\infty} A^n B^n \exp(-nsT) \qquad (1)$$

Now suppose that the signal pulse transmitted in this system has the Laplace transform H_1 , then a signal pulse delayed by a time $n_1 t_1$ has the transform $H_1 \exp(-sn_1 t_1)$. A sequence of such pulses occurring at multiples of time t_1 has the transform :-

$$S_1 = H_1(1 + \exp(-sn_1 t_1) + \exp(-sn_2 t_1) + \exp(-sn_3 t_1)\ldots)$$

where $n_1, n_2, n_3 \ldots$ are integral. (2)

In the special case where the pulses form an evenly spaced train, $n_1, n_2, n_3, \ldots = 1, 2, 3, \ldots (N-1)$

then $$S_s = H_1 \frac{(1 - \exp(-sNt_1))}{(1 - \exp(-st_1))} \qquad (4)$$

Now if we apply the sequence of pulses having the transform S_1 to the filter having the response V_o/V_1 given above, the transform of the output is :-

$$S_o = S_1 \times \frac{V_o}{V_1} \qquad (5)$$

$$= H_1(1 + \exp(-sn_1 t_1) + \exp(-sn_2 t_1) + \exp(-sn_3 t_1)\ldots)$$
$$\times A \exp(-sT)\,(1 + AB \exp(-sT) + A^2B^2 \exp(-2sT)\ldots) \qquad (6)$$

We can match the delay T to the period t_1.

In the special case where $n_1=1, n_2=2, n_3=3$ etc., this gives :-

$$S_o = H_1 A\left\{\exp(-sT) + (1+AB)\exp(-2sT) + (1+AB+A^2B^2)\exp(-3sT) + \ldots\right\} \quad (7)$$

and in this special case, the inverse transform is :-

$$V_o = L^{-1}H_1A\exp(-sT) + L^{-1}H_1A(1+AB)\exp(-2sT) + L^{-1}H_1A(1+AB+A^2B^2)\exp(-3sT) + \ldots, \quad (8)$$

Thus in this special case the output is expressed as a train of pulses spaced at intervals of T.

It is of interest to note that in the special case where $AB = -1$ (i.e. negative feedback, unity loop gain) :-

$$V_o = L^{-1}H_1A\exp(-sT) + L^{-1}H_1A\exp(-3sT) + L^{-1}H_1A\exp(-5sT) + \ldots\ldots\ldots \quad (9)$$

Thus in this special case, every other pulse is cancelled out by the preceding pulse, i.e. such an arrangement would perform the function of pulse frequency halving.

We are interested in the case where the loop gain AB is positive. What is the condition for the pulses to continue at constant amplitude after an input pulse train has finished ? We are particularly interested in the condition for a single pulse to be retained.

If we insert a single pulse having transform H_1, then the output transform is :-

$$S_o = H_1 A\exp(-sT) + H_1 A^2 B\exp(-2sT) + H_1 A^3 B^2 \exp(-3sT) \ldots$$

Now if we require the pulse to be maintained recirculating with unchanged amplitude, we must have :- $1 = AB = A^2B^2 = A^3B^3 = \ldots\ldots$

This is fulfilled by the conditions $A = 1$ and $B = 1$. However, if $A > 1$ we simply have an amplification of the input pulse amplitude by a factor A.

If we make the loop gain AB slightly less than unity, then the stored pulses gradually decay exponentially, the N th output pulse being $A^N B^N$ times the amplitude of the first output pulse.

The decay time constant is determined by :-

$$A^N B^N = 0.368$$

whence
$$N = \frac{\log(0.368)}{\log(AB)}$$

Alternatively, the required loop gain can be determined from :-

$$AB = (0.368)^{1/N}$$

A loop gain of 0.99 will cause the amplitude of the 100 th recirculated pulse to be 0.368 of the amplitude of the first pulse.

The curve of fig. 11.1 shows the required loop gain AB plotted against the number of pulses recirculated before decay to 0.368 of the amplitude of the first pulse has occurred. A partial table corresponding to this curve is given below.

N	AB	N	AB
2	0.6059	6	0.8467
3	0.7166	10	0.9048
4	0.7789	100	0.9901
5	0.8189	1000	0.9988

From these results, it is seen that a very close control of the effective loop pulse gain is required if the remembered pulses are to decay only slowly.

The effect of noise on such a recirculating storage loop was considered in sec. 7.5. It is of interest here to consider the ratio of output noise to input noise as it varies with the value of N. The result is plotted in fig. 11.2. It is seen that the use of positive feedback in a linear system is not a desirable form of memory if the decay is required to be slow.

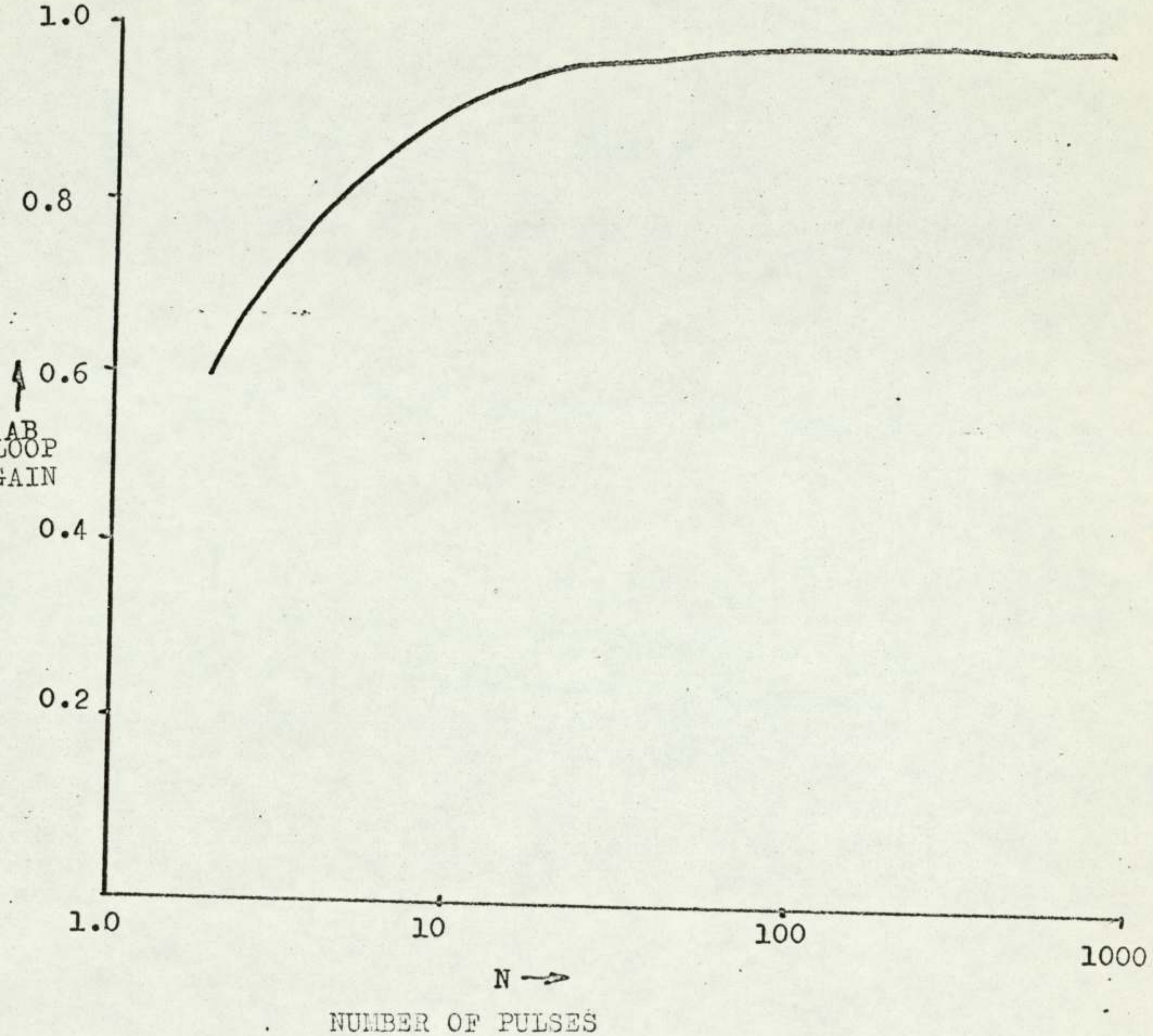


fig. 11.1

LOOP GAIN AS FUNCTION OF NUMBER N OF PULSES

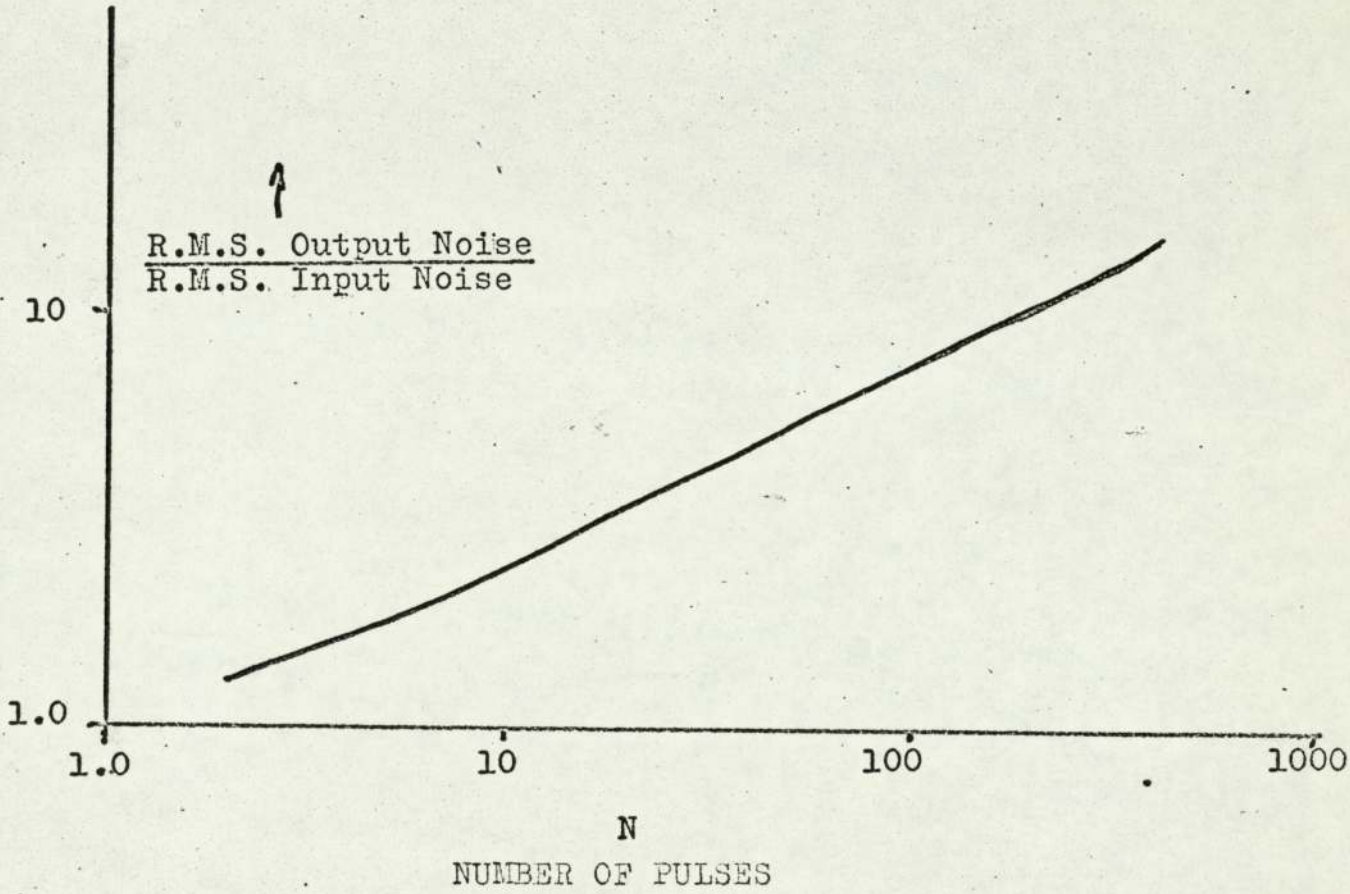


fig. 11.2

NOISE RATIO AS FUNCTION OF NUMBER N OF PULSES

From the foregoing it is seen that there are possibilities of simulating the hypothetical neural method of short-term storage of information by making use of recirculating positive feedback electronic loops involving a delay, and by storing short pulses. However it is undesirable to require that a separate recirculating loop be required to store each separate item, since this would lead to excessively complex memory arrangements for storage of quite small amounts of association information.

A single recirculating loop can be used to store a number of different items provided that steps are taken to avoid mutual confusion. It is possible to base the design of such arrangements on the use of the properties of prime numbers.

Any two prime numbers have an exclusive product, i.e. no other two prime numbers can have the same product. This fact follows directly from the basic properties of primes. The possibility of utilising this fact was considered with the earlier sinusoidal storage scheme, but the nature of the logarithm of a prime number introduced difficulties. However, in the present instance use can be made of the prime numbers themselves rather than of their logarithms, as will be seen later.

11.2 NON-LINEARITY

It will be noticed that in the above derivations, no specific steps have been taken to introduce non-linearity into the loop gain AB. Biological nervous transmission appears to include continuous pulse re-shaping. Similar pulse reshaping processes would be very easy to introduce into electrical storage loops, for example by the addition of a single saturating-type of non-linearity at one point in the loop.

However, such a storage loop could not possess the feature of slow decay of stored information which was required in the present investigation. In a saturating type loop, the gain around most of the loop would exceed unity, and the saturation non-linearity would then ensure that the overall effective gain was unity. There is no known way in such a non-linear loop of providing an overall effective gain very slightly less than unity in order to ensure the required slow decay of the stored information.

One possibility here is the progressive narrowing of stored pulses in the loop. This might be done for example by slightly delaying the leading edge of a pulse while not delaying the trailing edge. While this would ensure the slow decay of stored information, it would necessitate circuit complication. It was decided instead to limit the investigation at this atage to a considerat-ion of linear loop gain.

12. POSSIBILITY OF NON-DYNAMIC RECIRCULATION

12 POSSIBILITY OF NON-DYNAMIC RECIRCULATION

Following the precedent set by the apparent method of operation of the short-term memory arrangements in biological systems, the work described earlier was all based on the use of dynamic storage of information. Such an approach shows some promise for use in engineering systems. However, it has been shown that difficult problems remain to be solved in the practical development of such dynamic systems.

The problems are mainly those, such as noise augmentation, introduced by the dynamic nature of the storage arrangements. If dynamic storage is indeed used in biological information storage systems, it is clear that these problems can be overcome. However, this fact does not necessarily imply that the best engineering solution is the closest possible copy of the biological method.

It has been shown in the earlier work that a possible advantage of dynamic recirculation is the easy way in which the important forgetting or "information-overload-avoiding" process could be incorporated automatically. This again is an advantage if a direct simulation of biological processes is required. However there are other possible ways in which a forgetting process could be incorporated, and it might be better to consider the use of these alternative methods in engineering systems. If the alternative methods are adopted, then the absolute necessity for the use of dynamic storage methods no longer exists, and the use of non-dynamic methods becomes worthy of investigation. Such non-dynamic methods, together with forgetting arrangements, are required anyway for permanent cybernetic memory systems.

A dynamic memory arrangement loses its entire content of stored information in the event of even temporary failure, for example of the power supply circuits. While biological evidence apparently indicates that this phenomenon does occur, for example in the case of concussion, it is only the information in the short-term memory which is lost and long-term memories do not disappear permanently. Thus it is very unlikely that long-term memories are stored in dynamic form.

There is a number of possible methods of non-dynamic storage of information. These can be divided into two general classes, namely :-

1. Those static methods in which all stored information is immediately available. An example is the use of magnetic cores.

2. Those methods in which any particular item of stored information is only available periodically. An example is the use of a magnetic tape or drum.

In the present work, immediate access is not essential, and so it is possible to concentrate on the second class of non-dynamic storage methods. Such a course has the advantage that the earlier work on dynamic storage systems made use of this class of storage methods. The experience gained and the equipment used in the earlier work was used and adapted for use in the present work. The possible use of magnetic cores in static storage systems is considered later.

When the magnetic tape or drum method of non-dynamic recirculation of stored information is to be used, the association signals are to be permanently recorded whenever an association occurs between inputs. The recording is then to be mechanically recirculated over and over again. This recorded association information can then be periodically extracted and used in the future if any of the previously-associated inputs re-occur. For example, the association-signals can be permanently recorded in pulse form at definite points on a continuous loop of magnetic tape. The recorded signals can then be continuously recirculated past a replay head by driving the loop of tape at a constant speed. A magnetic recording drum could be used in the same way. It should be noted that the recorded information is retained in magnetic form even if the power supplies are removed and if the mechanical drive ceases.

With a non-dynamic recirculation system such as that discussed, the information stored is determined entirely by the position on the magnetic medium at which a memory pulse is recorded. Consequently, any variation in speed of the mechanical drive arrangements of the magnetic recorder can have the effect of completely changing the apparent nature of the stored information. This could be a serious practical disadvantage of the system.

There are two possible solutions to this problem. First, the speed of rotation of the mechanical drive of the magnetic recorder could be rigidly controlled with reference to an extremely stable master frequency source. For example, the drive could take the form of a servo system in which the reference frequency was derived from a temperature-controlled quartz crystal oscillator of extreme stability. Such an arrangement would be quite complex.

Fortunately, there is a simpler possible alternative arrangement. In this method, all of the signals used in the overall system are derived from a single master source of clock pulses. The required master clock pulses are derived from a continuous train which is recorded on one "Master" track of the recording medium. Signal pulses having the various required repetition periods can be derived from this "master" track by counting, using for example electronic counters. Coincidences between the signal pulses can then be recorded on to a separate "Memory" track on the same recording medium as is used for the permanent master track. There is then no possibility of loss of synchronism between the source of master pulses and the source of memory signals, since both sources exist side-by-side on the same medium.

The two methods suggested here for overcoming the effects of variation of speed of the recording medium can be compared directly with the corresponding methods introduced by Williams and West[53] and by Booth[54] for use in the magnetic memories adopted in digital computers.

13. POSSIBLE USE OF CORE-STORES

13 POSSIBLE USE OF CORE-STORES

The method of information storage using magnetic cores is widely used in computers, and since it is completely static it should be considered for the present use. However, there are several problems. It would be necessary to have a large degree of redundancy in order to introduce an element of probability into the operation of a core store. Unfortunately, large core stores are expensive, and the extension of memory size is a quadratically increasing problem.

To handle 1000 inputs, it would be necessary to use 500,000 cores, even if considerations of redundancy are ignored. A tape store would have to operate with high-frequency signals to be capable of use with a large number of inputs, but this is not impossible as witness the development of video tape recorders. The inherent rapid access feature of core-stores is not a requirement in the present work.

Redundancy would be necessary because cores are essentially two-state devices. On the other hand, input signal summation and inhibition are not difficult features to add, and magnetic cores inherently carry out a pulse integral summation process to saturation. Economics rule out core stores for use at present.

Steinbuch has used core stores in his Lernmatrix. It is perhaps of interest here to list some of the criteria upon which he based this choice of storage medium:-

1. Input and output must be parallel (as opposed to serial).
2. No external supply must be required to maintain the information stored.
3. Newly-learned information must not affect stored information.
4. Read-out of stored information must be non-destructive.

The last three of these correspond to the requirements inherently adopted from the start of the present work. However, the first of these requirements does not correspond to any of the requirements adopted for the present work since it was accepted that some time delay in operation is not objectionable provided that it is kept short.

The important additional requirement considered in the present work is that the action should be probabilistic. While the present approach can be considered to embrace that of Steinbuch, the reverse is not true. One form of the Lernmatrix is non-binary in the sense that the input signals can be continuous.

14. USE OF SHORT PULSES

14.1 PRIME PERIOD STORAGE PRINCIPLE[38]

The basic idea behind the method of association memory storage using the properties of primes can be stated quite simply. The periods (rather than the frequencies as in earlier schemes) of the input signals are to be based on prime numbers. The repetition periods of input pulses are all to be different prime integral multiples of the period of a master clock pulse. Consider, for example, short pulses having a prime repetition period of 3 units of time, which occur during the same interval as other short pulses having a prime repetition period of 5 units of time. Assume that the two sets of pulses are applied to the two inputs of a logical AND gate. Then the output of the AND gate will consist of pulses of coincidence, having a period of 15 units of time. Such a signal is absolutely exclusive to the two prime repetition periods 3 and 5 . It cannot be produced by signals having any other prime repetition periods.

The method of use of this approach can be described with reference to fig.14.1. Here, pulses A, having a prime repetition period of three units of time are produced during the same interval as are pulses B, having a prime repetition period of five units of time. The two pulses therefore coincide every fifteen units of time. Pulses having this repetition period can be recorded to

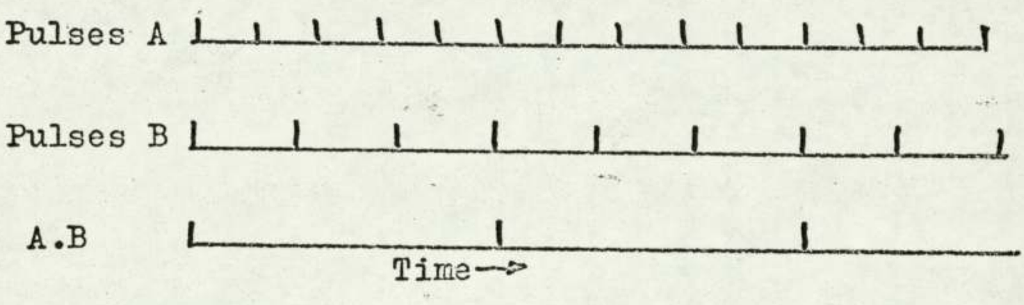


fig. 14.1

ILLUSTRATION OF PRIME PERIOD STORAGE METHOD indicate the past occurrence of coincidence.

If then in the future pulses A occur alone, they will coincide with the recorded information at every fifteenth interval, and this point in time also coincides with a B pulse. The coincidence of the new A pulses with the recorded A.B pulses can therefore be used to gate pulses from the B source to an output actuator if required. There can be no confusion with such a system because a pulse recorded at a given product interval can correspond only to two prime input intervals.
With this scheme, it is possible to interleave the storage so that a single storage medium can be used to store the associations between numbers of different prime input pulses. Such a feature leads to an economy of storage, an advantage shared by the earlier schemes using sinusoidal and using rectangular wave signals.

In order to use this scheme, sources of the various pulses are required, and so are coincidence detectors. If the coincidence detectors take the form of AND gates, then a large number of gates is required. For example, with five different possible pulse inputs A,B,C,D,E, all having different prime repetition periods, it is necessary to have ten AND gates for A.B ; A.C ; A.D ; A.E ; B.C ; B.D ; B.E ; C.D ; C.E ; D.E . Fortunately, matrix boards into which diodes can be plugged are now readily available, and use has been made of these in practical work on such systems.

14.2 BASIS OF THE SHORT-PULSE METHOD

A practical short pulse can be considered as the sum of a positive exponential rise and a delayed negative exponential rise. Thus :-

$$\mathcal{L}(\text{Short pulse}) = \frac{1}{s} \times \frac{1}{1 + sT} - \frac{1}{s} \times \frac{1}{1 + sT} \times \exp(-s\tau) \quad (1)$$

$$= \frac{1 - \exp(-s\tau)}{s} \times \frac{1}{1 + sT} \quad (2)$$

where T is the rise time constant and τ is the time length of the pulse. It has been assumed that the rise and fall time constants of the pulse are approximately equal in value.

In the present system, pulses like this are delayed by various times N_1t_a, N_2t_a, N_3t_a, etc., where t_a is the master pulse repetition period and N_1, N_2, N_3, etc. are numbers of the form :-

(Prime product)x(some number X)

Hence the Laplace transform of the recorded waveform is :-

$$\frac{1 - \exp(-s\tau)}{s} \times \frac{1}{1 + sT} \left\{ \exp(-N_1t_as) + \exp(-N_2t_as) + \exp(-N_3t_as) + \ldots\ldots \right\} \quad (3)$$

Now this recorded waveform is compared logically with inputs of the form :-

$$\frac{1 - \exp(-s\tau)}{s} \times \frac{1}{1 + sT} \times \exp(-N_xt_as) \quad (4)$$

where N_x takes the form (Prime number)x(some number Y)

If it ever occurs that

$$L \exp(-N_x t_a s) = L\left(\exp(-N_1 t_a s) + \exp(-N_2 t_a s) + \ldots\right) \quad (5)$$

where $$L = \frac{1 - \exp(-s\tau)}{s} \times \frac{1}{1 + sT}, \quad (6)$$

then an association corresponding to the present input is present on the recording and an output is required. The requirement is satisfied if ever

$$N_x t_a = N_y t_a$$ where N_y is one of N_1, N_2, N_3, etc.

or $$N_x = N_y$$

Now N_x and N_y are products of integers :-

N_x =(Prime number Q)x(Some number Y)

N_y =(Prime product RS)x(Some number X)

The equality is therefore satisfied if Q = R and Y=SX, or if Q = S and Y=RX. Note must also be taken of the case X = Q and Y=RS, since this places a limitation on the range of numbers which can be used in the system.

14.3 PULSE PRODUCT TRAINS

Consider an infinite series of pulses of unit height, each having a time duration of $2t_1$ seconds and with a repetition period of T seconds, as illustrated in fig. 14.2 below..

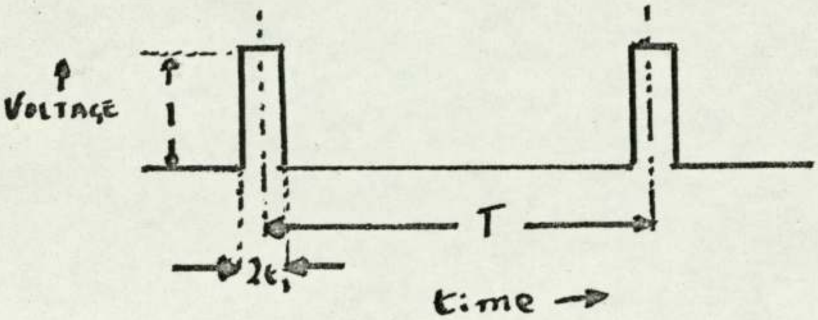


fig. 14.2

Such a pulse train can be expressed as :-

$$\theta(t) = \frac{2t_1}{T} + 2\sum_{n=1}^{n=\infty} \frac{\sin 2\pi n t_1/T}{n\pi} \cos \frac{2\pi n t}{T} \qquad (7)$$

However, if $t_1 \ll T$, then :-

$$\sin 2\pi n t_1/T \rightarrow 2\pi n t_1/T .$$

The amplitude of each harmonic then approaches $4t_1/T$.

For convenience, we can suppose that the pulse height is equal to $1/2t_1$, so that the amplitude of each component approaches 2/T, and equ.(7) becomes :-

$$= \frac{1}{T} + \frac{2}{T}\cos 2\pi\frac{t}{T} + \frac{2}{T}\cos 4\pi\frac{t}{T} + \frac{2}{T}\cos 6\pi\frac{t}{T} + \ldots \quad (8)$$

Now consider two separate pulse trains having pulse repetition periods of U and V respectively. The first train can be expressed as :-

$$\theta_1 = \frac{1}{U} + \frac{2}{U}\cos 2\pi\frac{t}{U} + \frac{2}{U}\cos 4\pi\frac{t}{U} + \frac{2}{U}\cos 6\pi\frac{t}{U} + \ldots \quad (9)$$

while the second train can be expressed as :-

$$\theta_2 = \frac{1}{V} + \frac{2}{V}\cos 2\pi\frac{t}{V} + \frac{2}{V}\cos 4\pi\frac{t}{V} + \frac{2}{V}\cos 6\pi\frac{t}{V} + \ldots \quad (10)$$

The product of these two pulse trains is therefore :-

$$\begin{aligned}\theta_1\theta_2 = \frac{1}{UV} &+ \frac{2}{UV}\Big(\cos 2\pi\frac{t}{U} + \cos 4\pi\frac{t}{U} + \cos 6\pi\frac{t}{U} + \ldots \\ &\quad + \cos 2\pi\frac{t}{V} + \cos 4\pi\frac{t}{V} + \cos 6\pi\frac{t}{V} + \ldots\Big) \\ &+ \frac{4}{UV}\Big(\cos 2\pi\frac{t}{U}\cos 2\pi\frac{t}{V} + \cos 4\pi\frac{t}{U}\cos 2\pi\frac{t}{V} + \\ &\quad + \cos 2\pi\frac{t}{U}\cos 4\pi\frac{t}{V} + \cos 4\pi\frac{t}{U}\cos 4\pi\frac{t}{V} + \\ &\quad + \cos 2\pi\frac{t}{U}\cos 6\pi\frac{t}{V} + \cos 4\pi\frac{t}{U}\cos 6\pi\frac{t}{V} + \\ &\quad + \ldots\ldots\Big) \end{aligned} \quad (11)$$

The final bracketed expression here can be expanded to give :-

$$\begin{aligned}\frac{2}{UV}\Big(&\cos\left(\frac{1}{U}+\frac{1}{V}\right)2\pi t + \cos\left(\frac{1}{U}-\frac{1}{V}\right)2\pi t + \cos\left(\frac{2}{U}+\frac{1}{V}\right)2\pi t + \ldots \\ &+ \cos\left(\frac{1}{U}+\frac{2}{V}\right)2\pi t + \cos\left(\frac{1}{U}-\frac{2}{V}\right)2\pi t + \cos\left(\frac{2}{U}+\frac{2}{V}\right)2\pi t + \ldots \\ &+ \cos\left(\frac{1}{U}+\frac{3}{V}\right)2\pi t + \cos\left(\frac{1}{U}-\frac{3}{V}\right)2\pi t + \cos\left(\frac{2}{U}+\frac{3}{V}\right)2\pi t + \ldots \\ &+ \ldots\ldots\Big)\end{aligned} \quad (12)$$

Each term of the expansion can be written in the form :-

$$\frac{2}{UV}\cos\left(\frac{a}{U} \pm \frac{b}{V}\right)2\pi t = \frac{2}{UV}\cos\left(\frac{aV \pm bU}{UV}\right)2\pi t$$

Now we know that the product wave must take the form of a single train of pulses having a pulse repetition period of UV. It can be expressed as :-

$$\theta_1\theta_2 = \frac{1}{2t_1} \times \left[\frac{1}{UV} + \frac{2}{UV}\left(\cos 2\pi\frac{t}{UV} + \cos 4\pi\frac{t}{UV} + \dots\right)\right] \quad (13)$$

It follows that in the expansion we must have

$aV \pm bU$ = every positive integer.

It is worth noting that if the original pulse trains are AC coupled to a true multiplying circuit so that a true product is obtained, the first two terms of the $\theta_1\theta_2$ product expansion of equation (11) disappear. The final bracketed terms must therefore contain a product pulse wave, together with other negative pulses if the two pulse trains are AC coupled to a true multiplying circuit. However, a logical AND circuit will merely give the positive coincidence pulses without the undesired negative pulses.

It has been assumed in the above discussion that the pulses are short. If this is not so, there can be partial coincidence between several successive pulses and a true product pulse train is not obtained.

14.4 PULSE-WIDTH TOLERANCE

If the pulses used are derived by counting and gating directly from a high-frequency rectangular-wave generator, then there should be no possibility of partial coincidence between several successive pulses from two sources having different prime repetition periods. However, the maximum usable frequency of the high-frequency rectangular wave generator will depend on the possibility of mutual interference.

Such spurious mutual interference could be caused, for example, by pulse lengthening caused by carrier storage in semiconductor elements or by the effects of stray capacitance. The difficulty is illustrated by the ringed cases in fig.14.3which shows two pulse trains having prime repetition periods of 13 and 17. It is seen that the pulse-width tolerance or the pulse-position tolerance is only equal to one half of the pulse width in this case. However, this is not a serious limitation, since there is no reason why shorter pulses should not be used, the master signal then not having a one-to-one mark to space ratio. With any given system this course will, however, have the result of lowering the storage capacity.

Pulses of a high carrier frequency might be used to help to overcome tape defects and also to increase the number of channels which can be carried by a given memory tape.

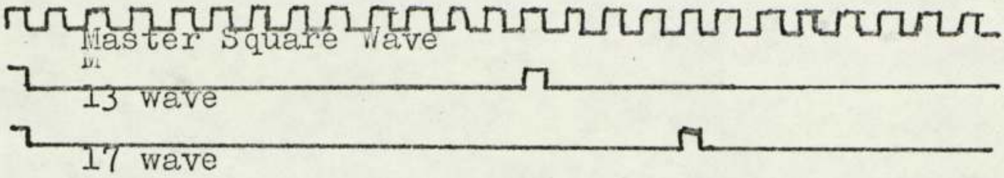
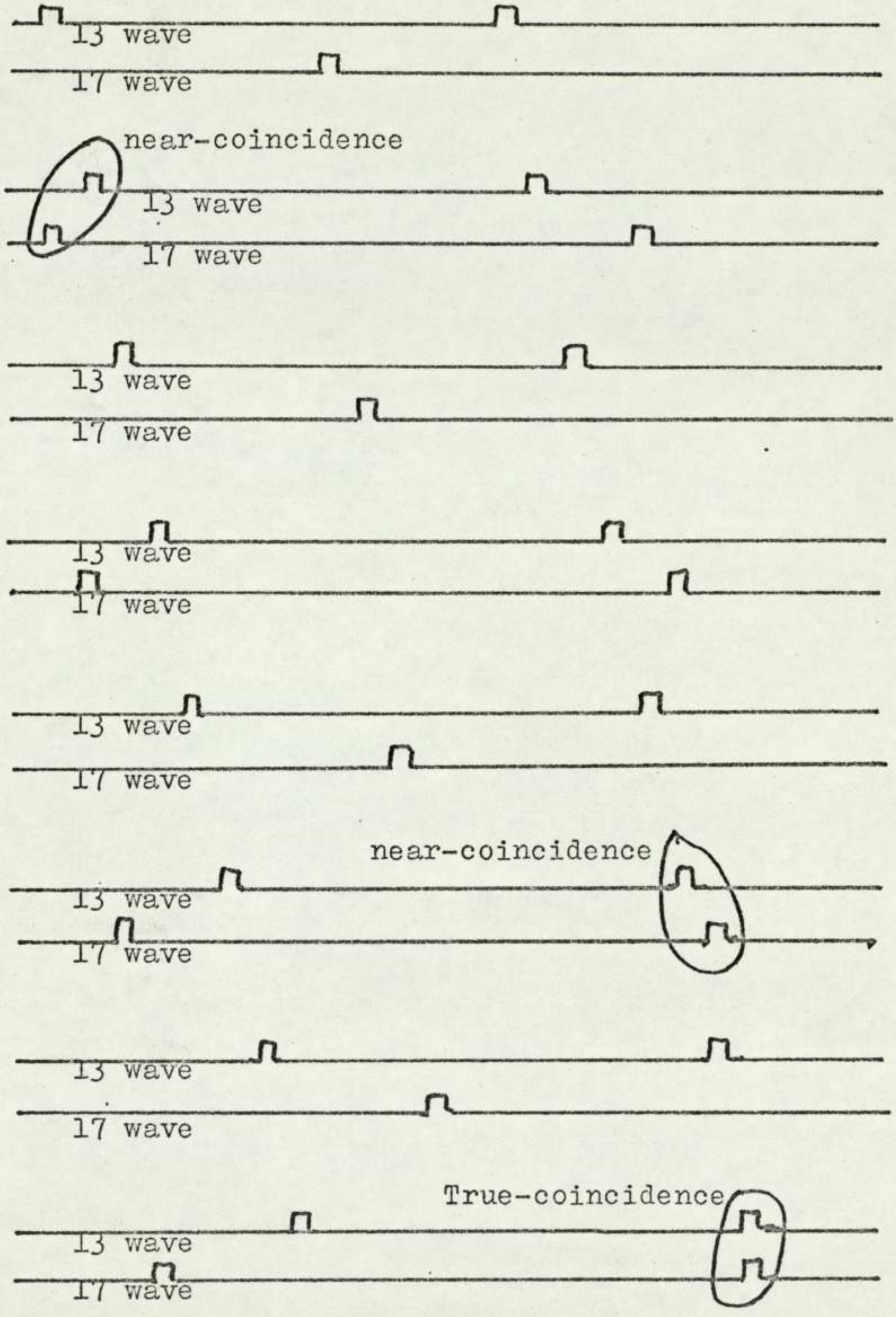


fig. 14.3 POSSIBILITY OF INTERFERENCE



14.5 LIMITATIONS ON RANGE OF PULSE PERIODS

It is not practicable to arrange a system to be capable of handling pulse periods as high as :- $t_a \times A \times B \times C \times \ldots \ldots \times X \times Y$, where A,B,C,.... are all different prime numbers and t_a is the master pulse repetition period, since the numbers involved would become astronomical. It is therefore necessary to impose some practical limitation on the maximum number of inputs which can be associated. Suppose that it is decided that only pair-associations are to be accepted. What are the limitations on the prime numbers which can be used if there is to be an inherent rejection of pulses produced by triple associations ? The rejection process is not difficult if

$$X.Y < A.B.C$$

where X and Y are the two longest prime pulse periods used (divided by the master pulse period t_a) and A,B,C, are the three shortest prime pulse periods used (again divided by t_a). In this case, the system is simply arranged to reject any pulses having a repetition period greater than $X.Y.t_a$.

The requirement $X.Y < A.B.C.$ can be approximated

by $$X^2 < A^3$$

or $$X < \sqrt{A^3}$$

i.e. the largest prime number must be less than the square root of the cube of the smallest prime number. This relationship is illustrated by the curve of fig.14.4 The range can be expressed by writing :-

$$X < RA$$

whence $R \approx \sqrt{A}$ gives the approximate ratio of the largest prime number used to the smallest prime number used. However, it is also necessary to consider the distribution of prime numbers in the range in order to find how the number of available channels varies.

Now the number of primes less than any given number N is approximately equal to $\frac{N}{\ln N}$, as is well known.[39] Consequently the number of primes between A and X is given by :-

$$\frac{X}{\ln X} - \frac{A}{\ln A} = \frac{A(\frac{2}{3}\sqrt{A} - 1)}{\ln A}$$

This number of primes between A and $\sqrt{A^3}$ is plotted against the value of the smallest prime number A in fig.14.5. On this basis, the number of available channels increases rapidly as the value of the smallest prime number to be used is increased.

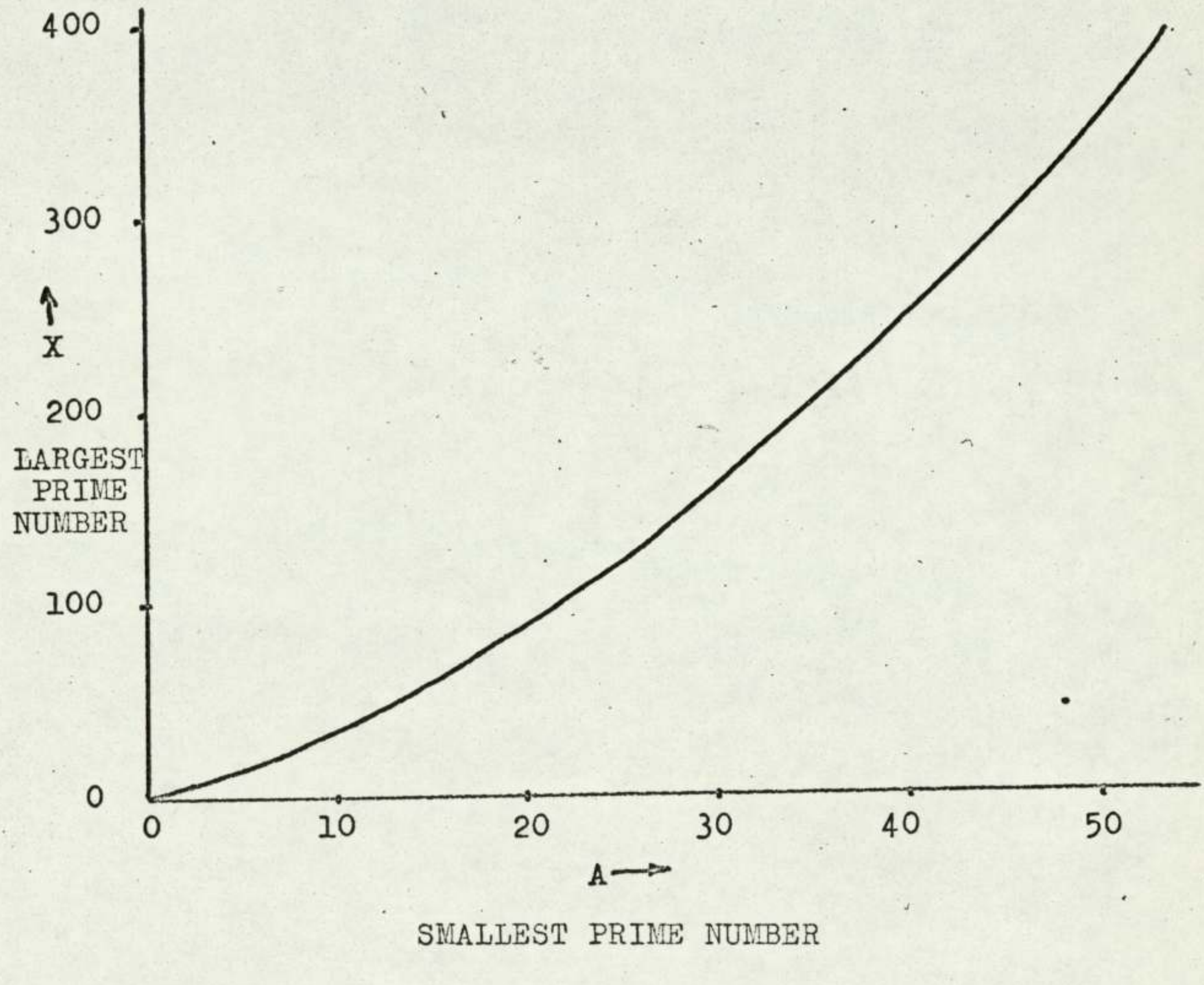
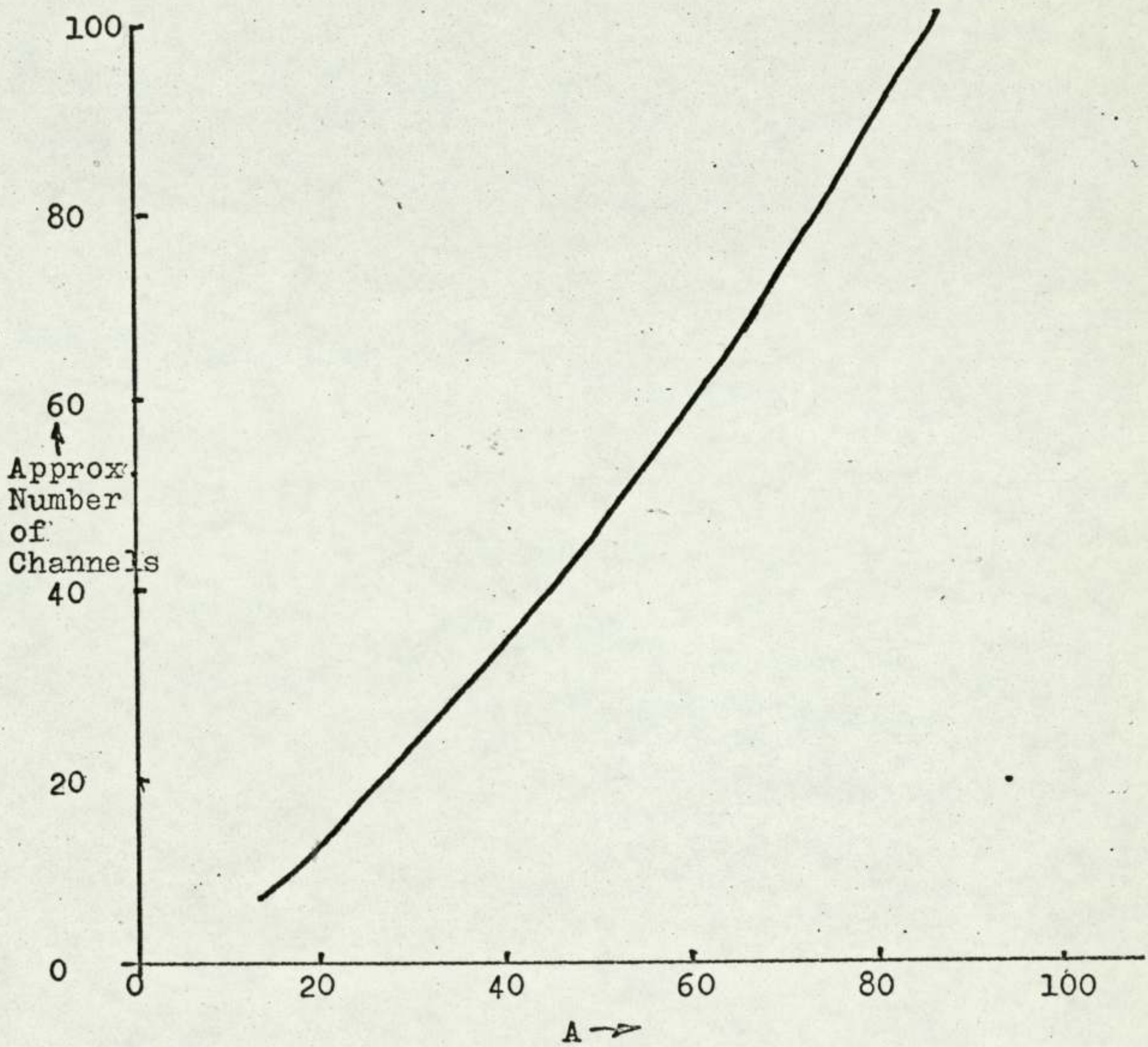


fig. 14.4

RELATIONSHIP BETWEEN LONGEST AND SHORTEST PERIODS



SMALLEST PRIME NUMBER

fig. 14.5

DEPENDENCE OF NUMBER OF CHANNELS ON SHORTEST PERIOD

14.6 AVOIDANCE OF SPURIOUS RESPONSE

If every coincidence of two stimulated prime number inputs is recorded, there is an additional reason for limiting the range of numbers used. Consider a coincidence of the two shortest prime input numbers used, A and B. A pulse is recorded at positions of the memory track corresponding to positions :-

AB, 2AB, 3AB,.............NAB

on the master track. Now N might be equal to one of the prime numbers A,B,C,... used in the system. Suppose that N = D, then NAB = DAB. A pulse recorded at this position would falsely indicate that input D had been associated with input A and with input B. This spurious response can be avoided if the ratio of the largest prime product XY to the smallest prime product AB is less than the smallest prime A, or :-

$$\frac{XY}{AB} < A$$

i.e. $$XY < A^2B$$

Approximately, this once again gives the design condition :-

$$X < \sqrt{A^3}$$

14.7 CONCLUSIONS

It has been seen that the achievment of electronic forms of neural propagation and storage of information in pulse form is quite feasible. In an engineered system, the use of prime pulse periods can give economy of storage fequirements, while avoiding the possibility of ambiguity of the stored associatian information.

With a recirculating pulse type of system, reshaping of the pulses can help to reduce the effects of cumulative distortion in the memory loop. A very close control of the effective recirculating memory loop gain is essential if it is to be ensured that the forgetting process is not to be excessively rapid. If alternative methods of forgetting can be introduced, non-dynamic methods of storage can be used, thus eliminating these difficulties.

Ambiguity problems can be avoided by the use of sufficiently narrow pulses. The necessity for precise control of memory tape speed can be avoided by the use of a master pulse track on the recorder as a source of all pulses. Prime-period counters are then required in order to generate the required prime-period pulses from the master track.

15. THE PRIME-PERIOD PULSE SYSTEM

38

15.1 A MACHINE USING THE PRIME-PERIOD PULSE SYSTEM

It was decided to construct a machine operating on the prime-period pulse system and using non-dynamic storage on a loop of magnetic tape. A block diagram of the system is shown in fig. 15.1.

Input stimuli are applied to the machine by means of switches P_1 to P_4. In order to record the association of two or more stimuli, the corresponding switches are closed coincidentally.

To achieve reliable recording of this information, the duration of the switch closure must obviously exceed the cycle time of the tape store.

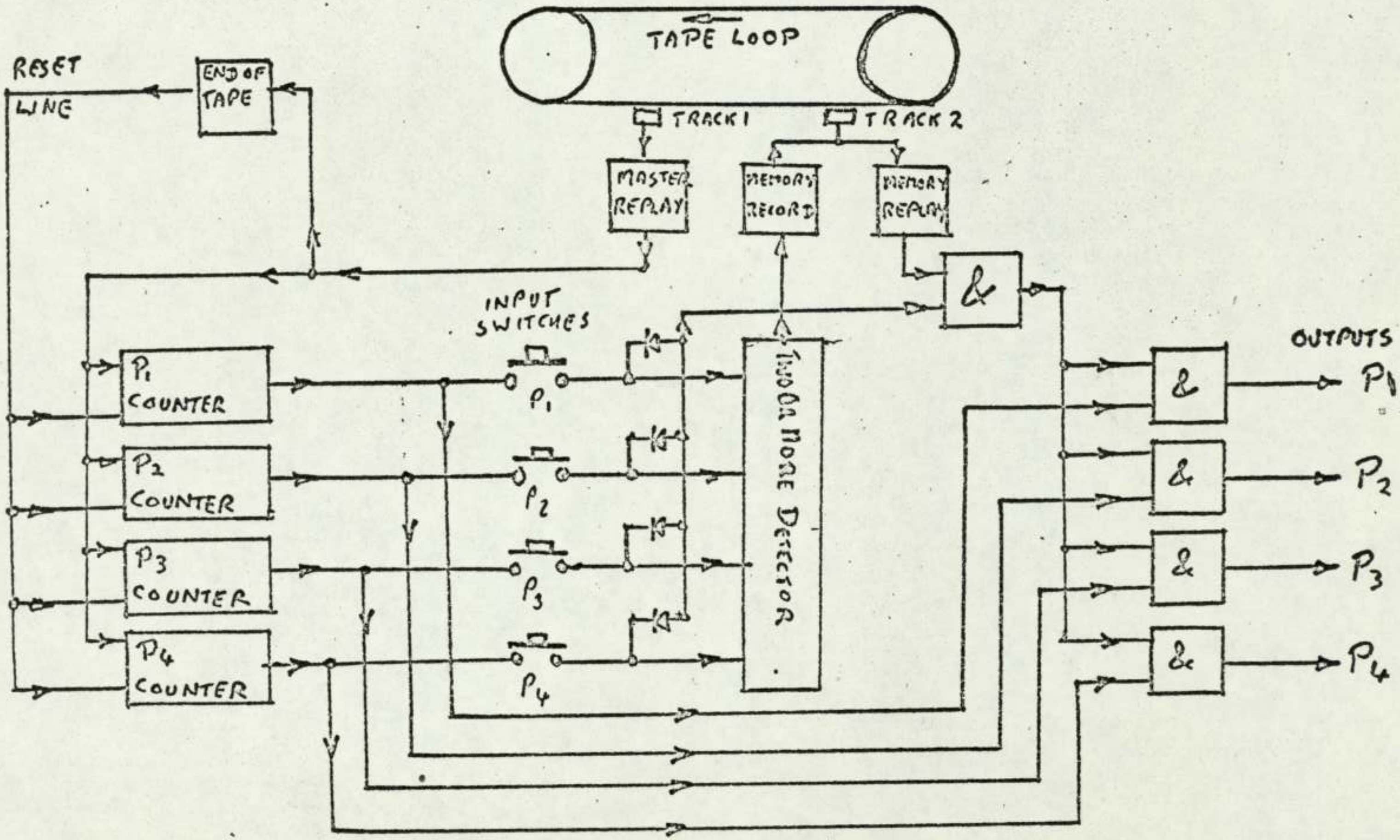

TAPE LOOP
TRACK 1
TRACK 2
RESET
LINE
END OF TAPE
MASTER REPLAY
MEMORY RECORD
MEMORY REPLAY
&
INPUT SWITCHES
P_1 COUNTER
P_2 COUNTER
P_3 COUNTER
P_4 COUNTER
P_1
P_2
P_3
P_4
TWO OR MORE DETECTOR
OUTPUTS
P_1
P_2
P_3
P_4

fig 15-1

15.2 PRIME PERIOD PULSE SYSTEM DESCRIPTION.

In the basic prime-period pulse system a loop of magnetic tape has two recording tracks. One of these (the Master Track) produces a continuous stream of equally-spaced short pulses from a replay head. After amplification and shaping these pulses are applied to a number of electronic counters as shown in fig. 15.1 Each of these counters is designed to count repetitively a certain number of master pulses and then to give an output pulse, reset itself to zero and then to restart counting. The number of master pulses required before any particular counter gives an output pulse is arranged to be a prime number. Thus the output pulse from one counter might coincide with every 17th master pulse, another with every 29th master pulse and so on.

The output pulses from the counters can be supplied via mechanical or electronic switches to a "Two-or-More" detector. This circuit is arranged to give an output pulse if, and only if, pulses are applied to at least two of its input terminals coincidentally. Thus if, for example, the switch connected to the "13's" counter is closed at the same time as the switch connected to the "17's" counter, then the "Two-or-More" detector produces an output pulse coincident with every 13 x 17 = 221st master pulse. Now since prime numbers produce unique products, this "221" pulse can be produced by no two other input switches than the "13" and the "17" switches. Further discussion is required if it is possible for more than two switches to be energised simultaneously, and this will be dealt with later.

Any output pulse from the "Two-or-More" detector is supplied to a recording circuit which records on the second, "Memory", track on the tape loop. In this way, if ever two or more inputs are energised simultaneously a unique record is made of this fact on the memory track on the tape loop.

Suppose that at some time in the past such a record has been made of two inputs, say "13" and "17", being energised simultaneously. A pulse has therefore been recorded at position 221 on the memory track, and this pulse is played back at the appropriate time each time that the correct piece of tape passes the record-playback head. Such played-back pulses are shaped and amplified and then passed to a logical AND gate. Another input to this AND gate is provided from the counters whenever one of the inputs is energised.

Now suppose, for example, that at some time in the past the "13" and the "17" inputs have been energised simultaneously, so that a pulse has been recorded on the memory track at position "221". If at the present point in time the "13" input alone is energised,

then it gives out a pulse at each of the positions
13, 26, 39,..............208, 221, 234, etc.
The seventeenth pulse, at position 221, will coincide with the previously recorded pulse at this position which is now coming from the memory replay circuit. Both of these pulses are applied to the AND gate, which consequently produces an output pulse coincident with position 221.

This coincidence pulse conveys the information that the "13" and the "17" input have been energised together at some time in the past, and that at least one of them is currently being energised. The coincidence pulse is applied to a common line connected to one input of each of a further set of AND gates. The other input of each of these AND gates is supplied with pulses from one of the prime-number counters. Consequently, one of these AND gates can produce an output pulse if a pulse appears on the common line at a time position corresponding to some multiple of 13; the next requires a pulse on the common line at a position corresponding to some multiple of 17; the next some multiple of 19 and so on.

Thus if a pulse appears on the common line at position 221, then both the "13" and the "17" AND gates produce an output pulse, even though only the "13" input is at present energised. In this way, the information stored on the Memory track of the recording medium about the past association of the "13" and the "17" inputs is used to produce the effect of the "17" input even though the "17" input is not at present energised.

The problems encountered in the development of a machine to operate on these principles, and the techniques adopted, are discussed in later sections.

Pulses from the master track on the magnetic recording medium are used to drive the prime-number counters producing the various prime pulses for the different input circuits. There must be a joint at some point on the tape or drum, since at some point the end of the master track recording meets the beginning of the same track. It is necessary to take steps to overcome spurious counting caused by this joint in the continuous loop of recording medium.

It should also be noted that the counters all have to count different prime numbers P_1, P_2, P_3, etc. repetitively. If all are started counting at the point where they have just produced a mutual coincidence pulse with all other counters, they will not all be in mutual coincidence again until a number $P_1 \times P_2 \times P_3 \times \ldots$ of master pulses have been produced. This would take a very long time even if very short pulse-repetition periods were used. It is not practicable or desirable to use an extremely long tape loop. Consequently it is required that all prime number counters are periodically reset to zero, and this is conveniently arranged to coincide in time with the joint in the continuous loop of recording medium. A special end-of-tape detector circuit is consequently used in order to produce the required zero-reset pulses for the counters when the joint appears.

15.3 FRACTIONAL TIME ERRORS

In order that there shall be no confusion between pulses, although a time error is both desirable and inescapable in the present application, this error must have a maximum value which is less than the minimum time between pulses.

Assume that the time variation of pulses is $\pm t$. Then the minimum time between pulses is $T_1 - 2t$, where T_1 is the normal pulse interval.

Also assume that the speed variation of the recording medium can be expressed as :-

$$v = V_1(1 + f(t))$$

The minimum distance between recorded pulses on the recording medium is then given by :-

$$d_m = (T_1 - 2t).V_1.(1 - f'(t))$$

where $f'(t)$ is the peak value of $f(t)$.

Now when these two pulses are reproduced, let the variation of velocity of the recording medium be expressed by :-

$$v = V_2(1 + g(t))$$

Then the minimum reproduced time interval T_2 between two pulses will be given by :-

$$T_{2m} = \frac{d_m}{V_2(1 + g'(t))}$$

where $g'(t)$ is the peak value of $g(t)$.

Substituting in the value of d_m :-

$$T_{2m} = (T_1 - 2t) \cdot \frac{V_1(1 - f'(t))}{V_2(1 + g'(t))}$$

It will be seen from this that if $f'(t)$ and $g'(t)$ are both small and if $V_1 \approx V_2$, then the maximum error per pulse is t, as would be expected. However, if this is not so, then there is an additional time error.

The maximum fractional time error can be expressed as :-

$$\frac{T_1 - T_{2m}}{T_1} = 1 - \frac{T_{2m}}{T_1} . \quad \text{This becomes :-}$$

$$1 - \frac{T_1 - 2t}{T_1} \cdot \frac{V_1(1 - f'(t))}{V_2(1 + g'(t))}$$

$$= \frac{T_1(V_2(1+g'(t)) - V_1(1-f'(t)) + 2tV_1(1-f'(t))}{T_1V_2(1+g'(t))}$$

It should be noted here that if $V_1 = V_2$, as can normally be expected, then :-

$$\text{Max. error} = \frac{T_1(f'(t) + g'(t)) + 2t(1+f'(t))}{T_1(1 + g'(t))} .$$

Furthermore, if $f'(t) = g'(t)$ in addition, then :-

$$\text{Max. error} = \frac{2f'(t)}{1 + f'(t)} + \frac{2t}{T_1} \cdot \frac{(1-f'(t))}{(1+f'(t))}$$

expressed as a fraction of the normal pulse interval.

This approximation can be written as :-

$$\text{Max.error} = \frac{2f'(t)}{1 + f'(t)} \cdot \left[1 - \frac{2t}{T_1}\right] + \frac{2t}{T_1}$$

For values of $f'(t)$ up to 0.1, the convenient approximation can be used :- $\frac{2f'(t)}{1 + f'(t)} \approx 1.8f'(t)$

In general, it is required that the maximum error is equal to some constant fraction of T_1, say kT_1. Hence:-

$$1 - \frac{T_1 - 2t}{T_1} \cdot \frac{V_1(1 - f'(t))}{V_2(1 + g'(t))} = k$$

From this, it follows that :-

$$t = \frac{T_1}{2}\left[1 - (1-k)\frac{V_2(1 + g'(t))}{V_1(1 - f'(t))}\right]$$

Thus, given the mechanical drive characteristics of the recording medium, it is possible to determine the required amount of time variation for a given error function k.

16. PULSE COUNTERS

16.1 THE REQUIREMENT FOR COUNTERS

In the non-dynamic circulation scheme investigated, a number of pulse counters are required. These are used to divide the input pulse frequency by prime numbers.

Large numbers of these counters will be required in the eventual completed cybernetic machine. Each of these counters will be set up to divide the input pulse repetition frequency by a different prime number.

There are thus two outstanding requirements which must be met by the forms of counter to be used. These are :-

1. The counters must be inexpensive, since numbers of them will be required.
2. The counters must be easy to set-up to count any required number.

The counter inputs are to be drawn from a master pulse source common to all counters. The prime-number output from each of the counters has to be used to drive a number of gates of various types.

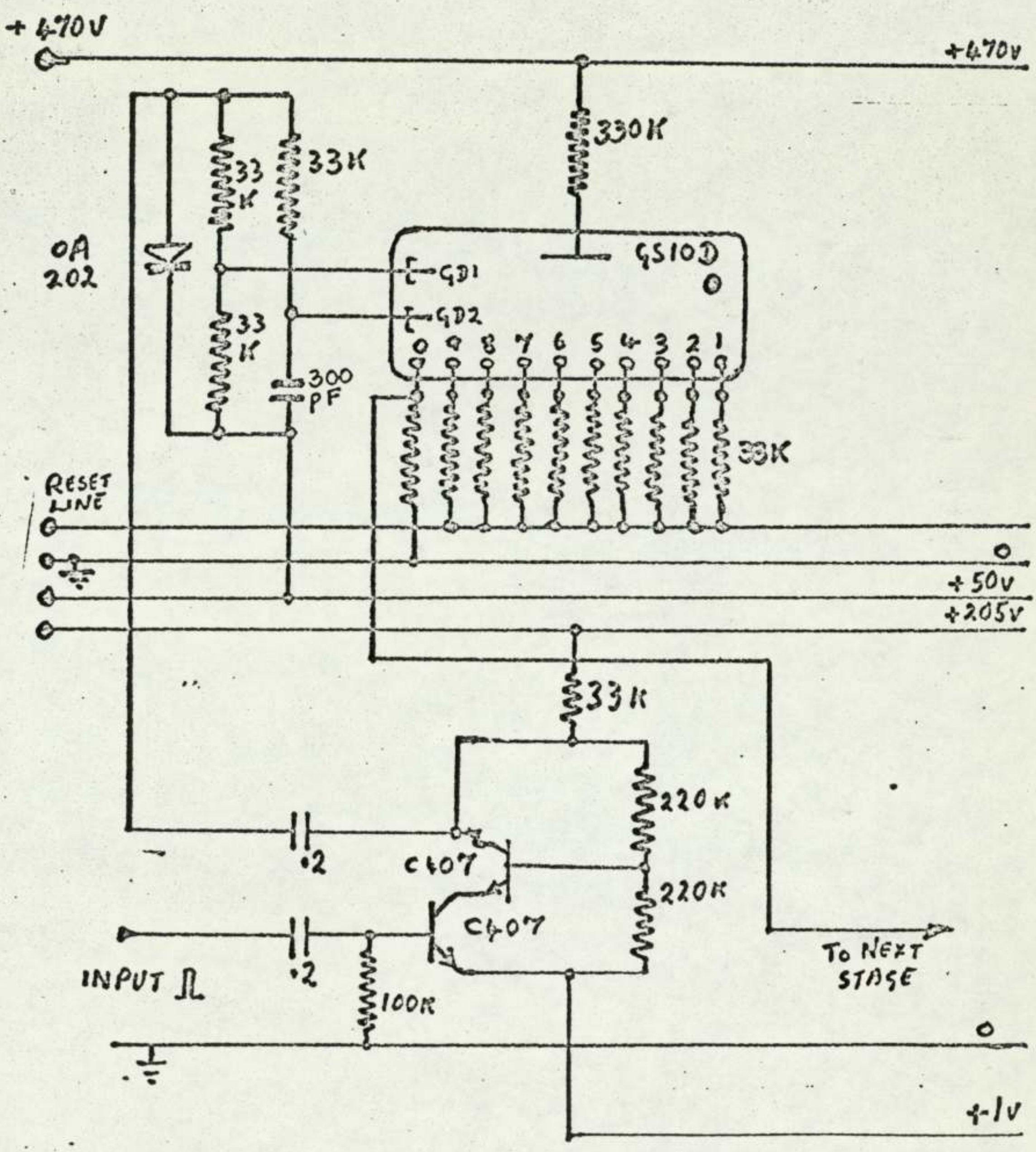
Counting techniques which were investigated are described in the sections below.

16.2 DEKATRON COUNTERS

The use of multi-output decimal or duodecimal counters such as trochotrons or dekatrons is attractive, since this would limit the number of inputs to AND gates required in deriving the prime-number sources. Trochotrons can be used to count at a comparatively fast rate, but they are expensive and their use would necessitate the provision of a heater supply.

Cold-cathode dekatron tubes claimed to be capable of a 50 kHz counting rate were available for a time. It was decided to avoid the use of these relatively untried tubes, and this proved later to have been the correct decision.

There was no reason why the initial work should not be carried out using a fairly low value of pulse repetition frequency. For example, if a 30 inch tape loop was operated at a speed of $7\frac{1}{2}$ inches per second, and the master track contained 10,000 pulses, then the pulse rate would be only 2,500 bits per second at a spacing of 333 bits per inch. If a 3:1 duty ratio is assumed, then the pulses would be 0.1 msec long, spaced by 0.4 msec. The counting of such pulses is within the capabilities of normal dekatron selector tubes.



ONE STAGE OF DEKATRON SELECTOR.

fig. 16.1

In the writer's earlier work using dekatron selectors,[40] thermionic valves were used for coupling. However, while the present research was in progress low-cost transistors having a reasonably high voltage rating became available. This made it possible to construct transistor circuits using the cascode principle and capable of providing the high-voltage pulses required for dekatron coupling.

The circuit diagram of one stage of the four-stage dekatron selector built for the present research is given in fig. 16.1. The writer has described the theory of operation of this coupling arrangement elsewhere[41] and it will not therefore be further discussed here.

If decimal counters were to be used in the application to a cybernetic associating machine, it would be necessary to arrange each counter to count to a different prime number and then to re-start its count. After the required prime number was counted, the counter would give an output pulse and also be reset to zero ready to re-start its count. This would require extra circuitry but it would not be difficult to accomplish. However, the chief disadvantage of the dekatron type of counter for the present application is large size. Also the constructional labour requirement is excessive, since large numbers of counters are required.

16.3 TRANSISTOR-DIODE INTEGRATED CIRCUITS

At the present time there are many different forms of integrated switching circuit which might be used in the construction of the types of device considered here. The integrated circuits which have been used in the present work are from a low-cost diode-transistor logic range which uses a negative supply rail in addition to the normal positive supply rail. It is perhaps worth mentioning at this stage the reasons for the selection of these devices.

The design and use of such circuits has received intensive investigation in the past with discrete component circuits.[42] The use of the diode-transistor form in industrial applications has shown its excellence for general-purpose use. It should perhaps be emphasized that in such general-purpose applications, unlike computer applications, the environment in which the devices must perform is relatively uncontrolled. This has introduced problems when attempts have been made to use integrated circuits, designed for computer use, in an industrial environment.

The use of diodes at the input to a unit can greatly reduce the problems encountered when units are interconnected. With some earlier forms of circuit, for example the resistor-transistor form, the performance obtained is very dependent on the output loading. With some other forms the available voltage swings are very small.

An important reason for the choice of the range selected is that in this range a negative bias supply rail is retained. Much attention has been devoted recently to the elimination of such rails in integrated circuits. It is perhaps useful to do so in digital computers, but for general purpose use the following points favour the retention of the negative bias supply rail :-

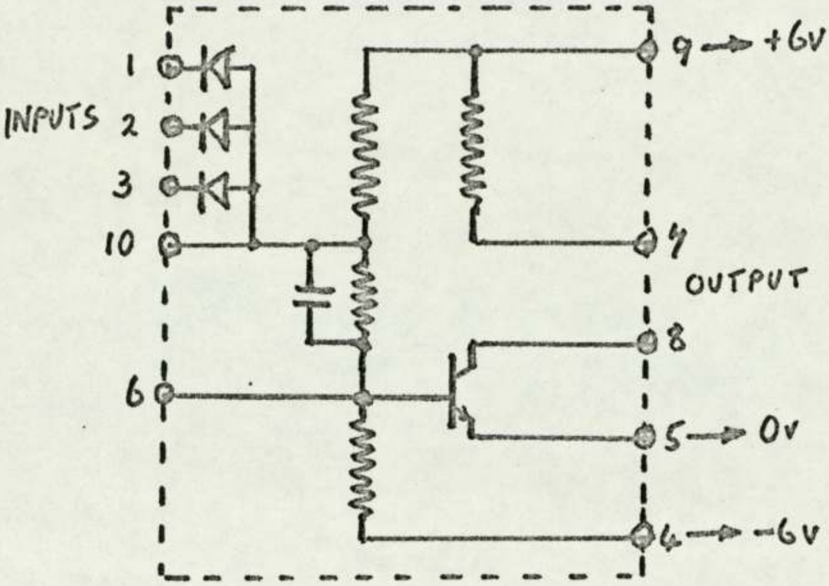
1. Improved noise immunity
2. Reduced temperature effects
3. Better pulse tolerance.

All of these points are a consequence of the fact that elimination of a bias supply rail makes it difficult or impossible ever to cut off a transistor properly by reverse-biasing the base. If any current is flowing in the forward direction into the base terminal of a transistor, then the device retains a current gain from base to collector. Consequently, slight variations of the base current caused by noise, temperature variations or by signal variations are greatly amplified in the collector circuit of the transistor. This is no longer true if the base terminal is truly reverse-biased so that reverse base current flows. The transistor then behaves merely as two reverse-biased diodes and there is no amplification of the minor effects, as long as the transistor is cut-off.

The bias supply rail which is required does not have to supply a heavy current or to have a very closely controlled value of voltage. Thus it is not an expensive feature to provide. Consequently, the diode-transistor circuit using a negative bias supply rail as well as a positive supply is well suited for the present application.

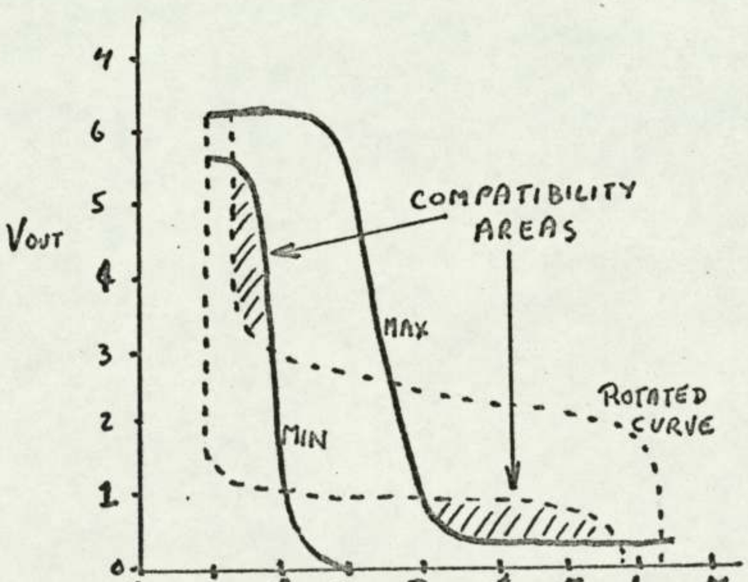
It should perhaps be mentioned that the range of integrated circuits adopted for the present work is no longer commercially available. Since it was the only known range which makes use of a negative bias supply rail, the devices will now be described briefly.

The circuit diagram of a typical unit of the type adopted, the OMY 120 NOR circuit, is shown in fig. 16.2.



OMY 120 NOR CIRCUIT

fig. 16.2



INPUT-OUTPUT VOLTAGE CURVES

fig. 16.3

The OMY range of integrated circuits was selected for use in the present research for the reasons outlined in the previous paragraph. The range uses a main supply voltage of +6 volts and a bias supply voltage of -6 volts. Diode-transistor logic is used. Physically, the TO5 can size is used, fitted with ten leads. The various types of circuit used will now be described briefly.

The OMY120 is a three-input NOR/NAND gate using an NPN transistor. In the writer's terminology [25] it is therefore a P1N circuit. Three input diodes are fitted, and their common connection is brought out so that additional external diodes can be fitted. It is a notable feature of this circuit that every component interconnection is brought out to one of the terminations. Consequently in the event of an open-circuit failure of a component, caused for example by accidentally exceeding ratings during experimental work, an external component can easily be wired in its place.

The circuit is capable of supplying an output current of 13.5 mA. The output impedance with a positive output is 2,900 ohms. This high value is immaterial with diode-transistor circuits, provided that the diodes at the inputs to the following stages have a high value of reverse resistance.

An input current of 2.75 mA must be allowed for, so for absolute safety each circuit output should not feed more than four inputs (this is sometimes stated as "fan-out = 4"). However, at room temperature it is safe to feed five inputs.

The input-voltage / output-voltage curves of this circuit come somewhere within a range which can be specified by two limiting curves. The usefulness of such circuits is determined by their ability to supply from their output terminals the inputs to other circuits. If two of these P1N circuits are connected together, with the output of each connected to the input of the other, a bistable circuit is formed. The output-input curves of the second circuit can be plotted on top of those for the first circuit as shown in fig.16.3, since $V_{o1} = V_{i2}$ and $V_{o2} = V_{i1}$. The extent of the bistability is determined by the overlap areas of the curves, and thus these overlap areas can be taken as a measure of the compatibility of the circuits.

With these circuits, interfering noise pulses of about 0.5 volts amplitude will not cause mal-operation. This is a valuable feature for use in experimental apparatus. On- and off- switching times of less than 0.1 microseconds are obtained, and this is adequate for the present application.

The OMY bistable integrated circuit has been used extensively in the present research, often in conjunction with the OMY122 steering circuit. Like the OMY120 gate, the OMY121 bistable circuit has the valuable experimental feature that every interconnection is brought out to an external terminal lead. This feature is unfortunately not shared by the OMY122 steering circuit. The OMY121 circuit is sometimes known as an R-S or Reset-Set flip-flop. It is capable of a counting speed of 2 MHz, which is ample for the present purposes. The compatibility and noise-immunity figures are similar to those for the OMY120 gate described above.

The OMY121 bistable circuit can be converted into a monostable circuit if one of the feedback paths is immobilised by connecting terminal 2 to the zero line. The delay time is then determined by the time constant of an added external CR coupling circuit. The circuit can be triggered by positive pulses to pin 3 or via an external coupling circuit by negative pulses to pin 6. This arrangement has proved to be very useful in the present research.

External components can also be added to the basic bistable circuit to convert it to an astable circuit or to an emitter-coupled voltage-sensitive trigger.

16.4 BINARY-DECIMAL COUNTING METHODS

The feedback binary-decimal method of counting previously described by the writer[43] was designed to overcome the spurious narrow "spike" waveforms usually encountered in such arrangements. These spikes occur when a stage which has just changed state is immediately changed back to its former state. In the method previously described, the spurious waveform was avoided by ensuring that a stage had always been settled down for a complete inter-pulse period before it was required to change state. Thus the state of the counter around the tens transfer point was :-

9	0000	
0	0001	
0	0111	(middle digits changed by feedback
1	1000	from outer two)

This arrangement has the following advantages :-

1. The middle digits have been steady for a complete inter-pulse period before they are changed by feedback.
2. The feedback pulse, which is also used for tens transfer, starts at a time when only the least significant digit is changing.
3. The feedback pulse is obtained on a definite output pulse rather than only while the output is changing.

4. The tens transfer pulse is also used for feedback, so that no special outputs are required.

5. At the time when the middle pulses are changing, they are not necessarily supplying output information since the first and last digits on their own give sufficient information.

These features make the method particularly attractive, and it appeared desirable to extend the system for use in the present prime-number counters. However, some thought about the system shows that it is unfortunately only suitable for use with even-number bases. The reason for this is that the transfer must always take place on the maintained occurrence of a particular value of the least significant digit. Consequently this method cannot be used for the present purpose, since the counters are required always to have an odd prime base. Many possibilities were considered. For example, one possibility is the use of counters always to count twice the required prime number, with a corresponding doubling of the input pulse repetition frequency. Another possibility is the acceptance of the "spike" waveforms encountered in many other counting systems.

If binary counters incorporating transistors are to be used, there is little sense in incorporating double decoding, first from binary to decimal and then to a prime number. Such a course would only have advantages if a visual indication of the count was required, since to most people a decimal form is easier for rapid understanding.

In the present case a visual indication is desirable for setting-up purposes, so it is useful to have visually-indicating decimal test counters in the equipment. However, it is not necessary for all of the counters in the prime-period pulse system to have visual indication, and universal incorporation of this feature would add unnecessarily to the complexity and expense of the system. Details of the commercial binary-decimal counters with visual indication which were used in this work are not given here, since the counters were later retained only for test purposes.

Unfortunately, no satisfactory flexible method of changing the counting base of a multi-stage binary counter was known in the early stages of the work, so it was necessary to devise such a method. Since large numbers of these counters were required, it was fortunate that at this point in the research low-cost integrated circuits became available.

16.5 POSSIBILITY OF COUNTING TWICE THE PRIME

One possible binary counting arrangement which was considered involved the detection of the condition when both the initial and the final digits first simultaneously achieved the value '1'. This condition can then be used to change some of the central digits from '0' to '1' in such a way that the count ends with all stages of the counter set to '1' after a total number of input pulses equal to 2N-1, where N is the required prime number. All stages of the counter will then attain the state zero simultaneously after a total number of input pulses equal to twice the required prime number N. The last few counts are then made equal to the complements of the first few, as required. The writer has discussed complementing binary codes elsewhere.[44] The method can best be illustrated by examples as shown in the table below.

This method has two disadvantages. First, an additional multiway AND gate is required with each counter to obtain the "all-zero's" output for recording. It is not sufficient merely to record when the first and last digits are simultaneously zero. Secondly, it is required to produce a pulse only at the point where the first and last digits are equal to 'one' for the first time. The necessity to ignore later coincidences involves circuit complexity.

1	00001			
2	00010			
3	00011			
4	00100			
5	00101			
6	00110			
7	00111			
8	01000			
9	01001			
10	01010			
11	01011			
12	01100			
13	01101			
14	01110			
15	01111			
16	10000			
17	10001	-	10111	
18	10010	-	11000	
19	10011	-	11001	
20	10100	-	11010	
21	10101	-	11011	
22	10110	-	11100	
23	10111	-	11101	
24	11000	-	11110	
25	11001	-	11111	
26	11010	-	00000	2x13

27	011011			
28	011100			
29	011101			
30	011110			
31	011111			
32	100000			
33	100001	-	111111	
34	100010	-	000000	2x17

33	100001	-	111011	
34	100010	-	111100	
35	100011	-	111101	
36	100100	-	111110	
37	100101	-	111111	
38	100110	-	000000	2x19

TWICE-PRIME COUNTING

The first time that the initial and final digits coincide with value 1, then some of the internal zeros are changed to ones as shown. The actual digits changed determine the prime.

16.6 PRIME-NUMBER COUNTERS USING INTEGRATED CIRCUITS

For the present application, the counters are required to be adaptable for repetitive counting of any odd-prime base. A possible method is based on the earlier approach,[43] but the new method[38] includes the use of a pulse delay of about one-half of the minimum interval between the master pulses which are to be counted. Such a delay is easily provided using an integrated monostable circuit; its use ensures the avoidance of "spike" waveforms which tend to produce unreliability in counters.

To illustrate the method, consider a four-stage integrated-circuit binary counter as shown in fig. 16.4 with feedback arrangements added to cause the count to repeat after 13 input pulses rather than after 16 input pulses. Each bistable circuit is associated with a steering circuit. The circuit arrangement of the individual stages is identical to that of the discrete component decimal counter mentioned in section 16.4.

In order to count 13, the feedback is required to perform the following binary switching process :-

10 1010

11 1011

12 1100

12 1111 (after delay)

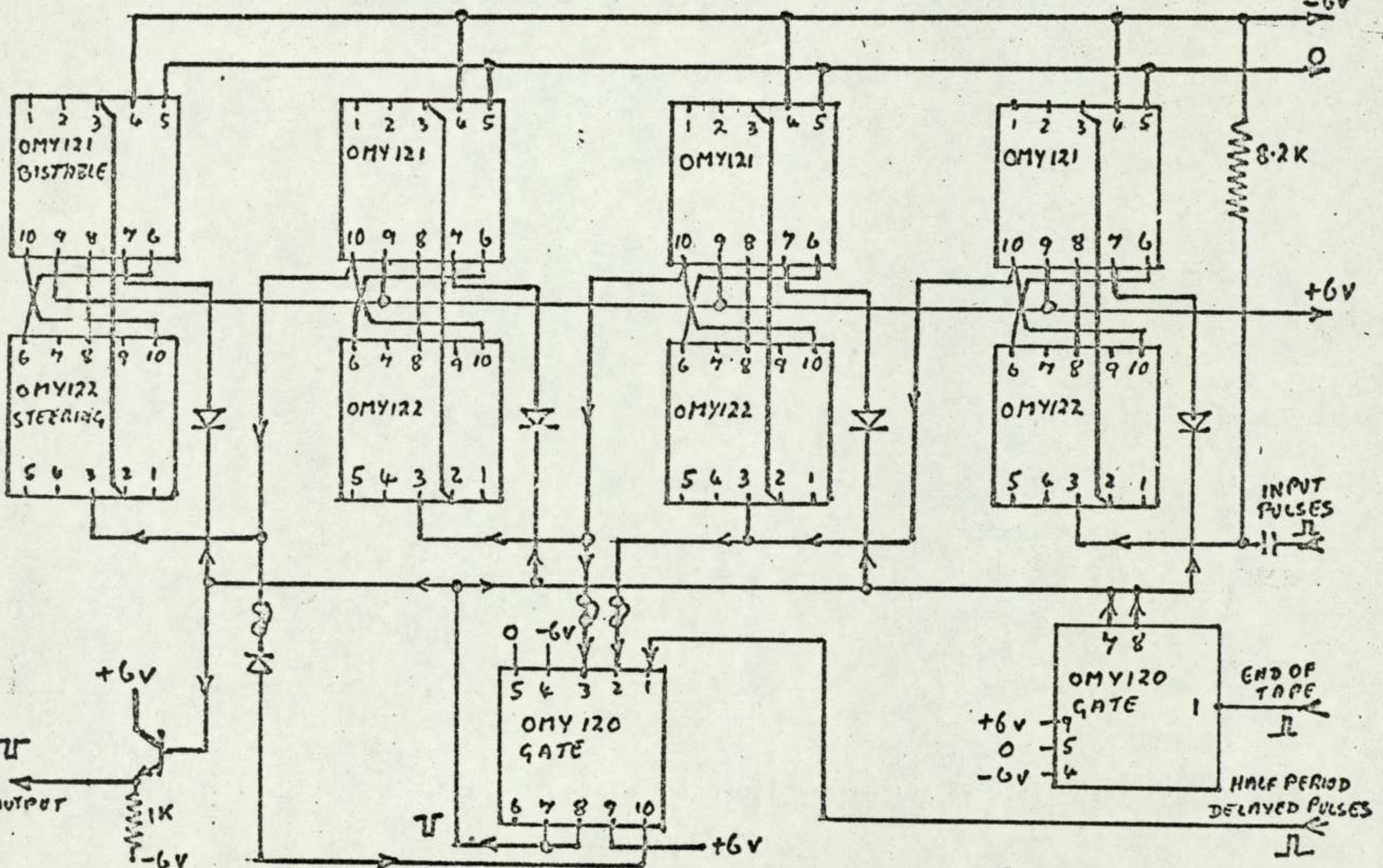
13(0) 0000

(1) 0001 and so on.

Thus, shortly after the counter indicates 1100 (i.e. 12), the state of the two least significant stages must be changed so that the counter indicates 'llll'. The next input pulse (the 13th) will then reset the counter to '0000'. The digit state is changed by a NOR circuit which resets the required bistable stages to the "one" state via diodes. In the case of the '13's' counter, the resetting is accomplished whenever both of the most significant digits have the value one for the first time, i.e. whenever the count is 12.

The master resetting pulse is taken to the NOR resetting circuit via a delay monostable multivibrator, (also using integrated circuits). In this way, resetting does not occur until all stages have settled down; the narrow "spike" waveforms previously mentioned are replaced by much longer pulses of about one-half of the standard length.

The length of the counting sequence, before resetting occurs, is determined by the inputs to the NOR resetting circuit. In practice all inputs to the NOR circuit, except the half-period delayed pulse, are taken to small sockets on the front of the counter unit, as are the outputs of the individual counter stages. In this way, the counting base is simply determined by the connections between the two sets of sockets. A standardised counter arrangement which can be set to count any required number base is thus achieved.



PARTIAL CIRCUIT DIAGRAM OF PRIME-NUMBER COUNTER

fig 16.4

16.7 HALF-PERIOD PULSE DELAY

It was decided to adopt the system described in section 16.6. This requires the use of a central circuit supplying pulses delayed by one-half period after the main pulse input to the counters. These delayed pulses are provided by an integrated monostable delay circuit, using the arrangement of fig. 16.5.

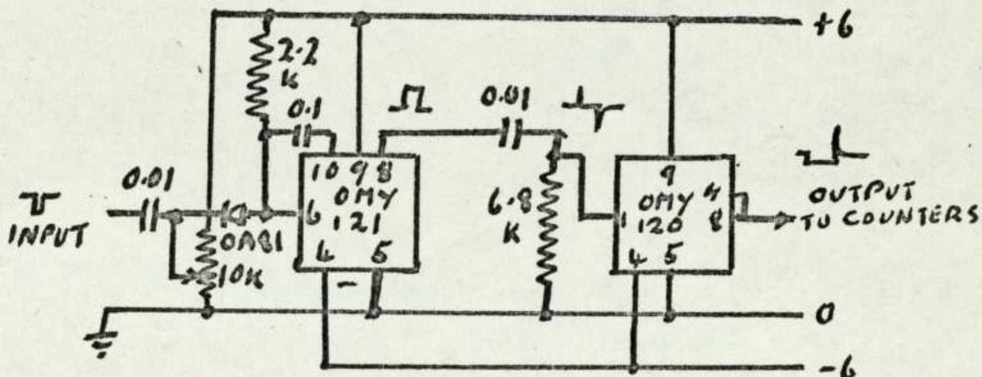


fig. 16.5 PULSE DELAY CIRCUIT

Negative input pulses from the master-track playback amplifier are used to trigger the monostable circuit, which then gives a positive pulse of length equal to about one-half of the master-track pulse repetition period. Each pulse is differentiated, and the trailing-edge pulse is amplified and phase-inverted before being applied to the reset circuits of all counters.

An integrated buffer-amplifier type OMY123 was originally used at the output. This is a bad design of integrated circuit which can apply a short circuit to the supply in certain circumstances, and so it was replaced by the OMY120 circuit as shown above.

17. MAJORITY LOGIC

17.1 MAJORITY LOGIC : GENERAL

One of the characteristic features of biological nervous systems is the logical property of neurons. The electrical output of some neurons appears to depend on the number of incoming dendrites stimulated simultaneously. It is important to note, however, the real difference between neural logic and electronic majority logic. Neural logic is continuous, in that the output, which takes the form of a pulse-frequency-modulated train, can convey any given point on a gradual scale of values. On the other hand, majority logic can convey only the two binary values "0" and "1".

It should be noted that because of this distinction, noise can only distort the output level of the signal in continuous neural logic : it cannot lead to an absolutely incorrect decision as it can in discrete logic. Perhaps it should be mentioned that the high-frequency noise components can easily be filtered out if a very rapid overall response is not required in electronic circuits.

A form of majority logic was required in the present work in order to detect when two or more inputs are simultaneously stimulated. When this occurs the two-or-more majority circuit must pass a pulse to the memory to record the occurrence of association between the two stimuli. The various circuits considered are discussed below.

17.2 RELAY MAJORITY LOGIC

Majority-logic circuits using electro-mechanical relays are well-known. An example devised by the writer for use in numerical-control equipment for machine tools[45] is shown in fig. 17.1 below. One of the output terminals is energised whenever an even number of input relays is energised. Circuits such as this can be regarded as degenerate forms of the complete relay contact circuit shown below in fig. 17.2.

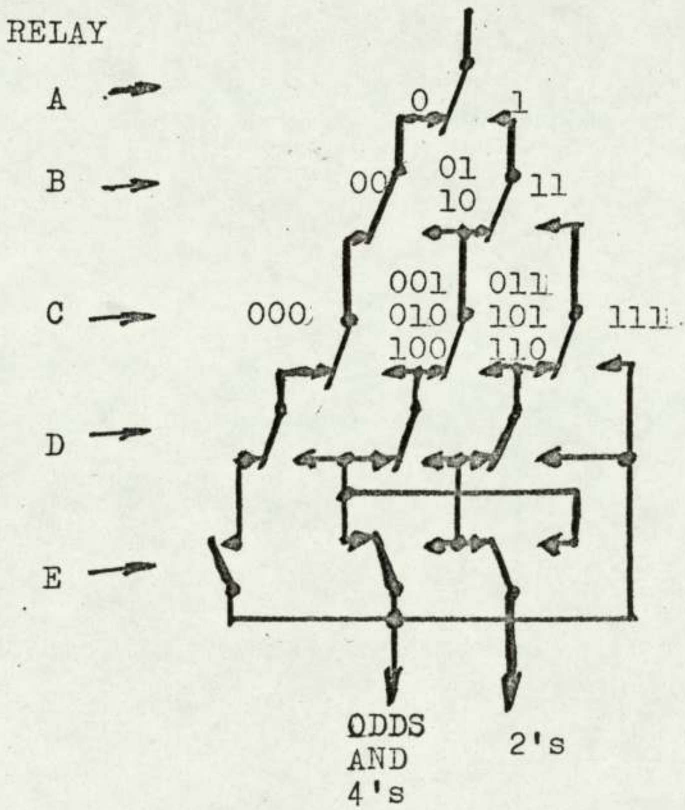


fig. 17.1 PUNCHED-TAPE PARITY CHECKING CIRCUIT FOR MACHINE-TOOL CONTROL.

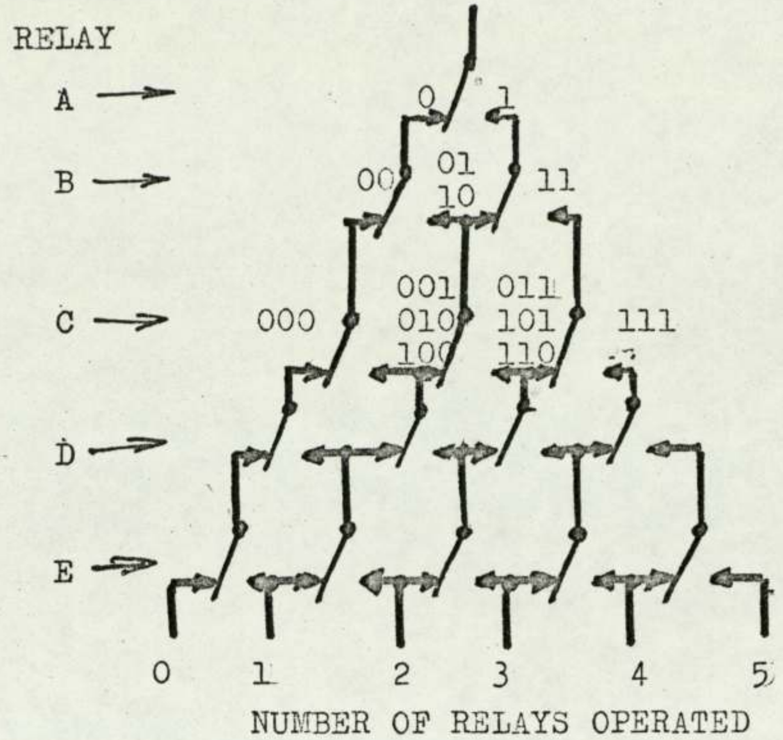


fig. 17.2 COMPLETE NUMBER-ROUTING CIRCUIT

A two-or-more circuit derived from the complete arrangement requires a maximum of two change-over contacts on each relay. A slight reduction of the contact requirements would be obtained by using complementary circuitry, i.e. by building a relay-contact "0 or 1" circuit and then inverting the output. The relay contact arrangement is shown for six relay inputs in fig. 17.3 below.

Solid-state relays could be used to increase the speed of action of these circuits. At present, this approach would be quite expensive. However it is hoped that integrated forms of these circuits will become available in the future.

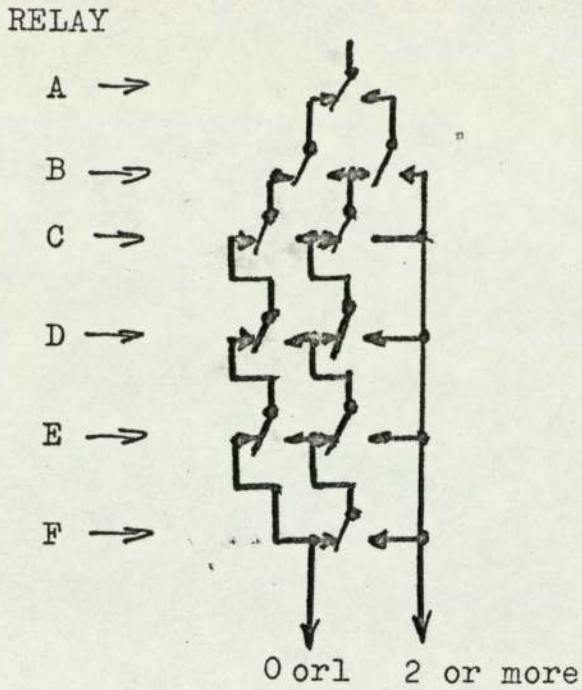


fig. 17.3 RELAY CONTACT TWO-OR-MORE CIRCUIT

After considering the problems involved in the use of such circuits with present-day techniques, it was decided to adopt other forms of two-or-more circuit.

17.3 RESISTIVE MAJORITY LOGIC.

There are two basic methods of achieving a circuit which gives an output signal whenever two or more input terminals are stimulated simultaneously. One uses binary logical AND gates. This approach is described in section 17.9 later.

The alternative approach requires less equipment, but there are difficulties in obtaining reliability. This method is sometimes called "Analogue Addition," and the basis of this method is illustrated by the circuit shown in fig. 17.4.

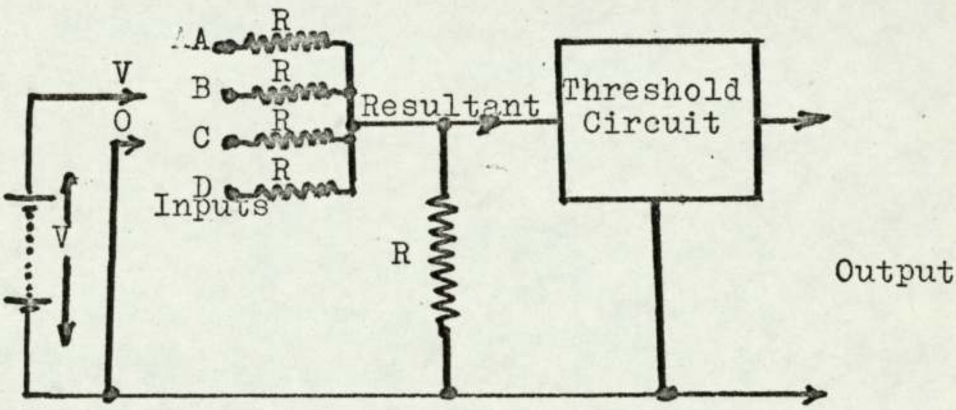


fig. 17.4

RESISTIVE MAJORITY LOGIC USING ANALOGUE ADDITION

Suppose that each of the inputs A,B,C,D in fig. 17.4 can be connected either to a voltage V or to a zero voltage O. The relationships between the input voltages to terminals A,B,C,D and the output voltage across the common shunt resistor are given in table 1 below:-

A	B	C	D	Output
O	O	O	O	O
O	O	O	V	0.2 V
O	O	V	O	0.2 V
O	O	V	V	0.4 V
O	V	O	O	0.2 V
O	V	O	V	0.4 V
O	V	V	O	0.4 V
O	V	V	V	0.6 V
V	O	O	O	0.2 V
V	O	O	V	0.4 V
V	O	V	O	0.4 V
V	O	V	V	0.6 V
V	V	O	O	0.4 V
V	V	O	V	0.6 V
V	V	V	O	0.6 V
V	V	V	V	0.8 V

TABLE 1

From Table 1 ,it can be seen that if the threshold circuit is designed to give an output only if the resultant voltage exceeds 0.3 V, then the output provides the logical indication of :-

$$V_o = A.B + A.C + A.D + B.C + B.D + C.D \qquad (1)$$

An output is therefore obtained only if two or more inputs are present.

For a general (N-1) input circuit of this type, the outputs are :-

Number of inputs stimulated	Output Voltage
0	0
1	V.1/N
2	V.2/N
3	V.3/N
......	....
N-1	V.(1 - 1/N)

Thus in order to detect the difference between the stimulation of X inputs and the stimulation of (X+1) inputs, the threshold circuit must differentiate between voltages which only differ by V/N. This becomes an increasingly difficult practical task as the value of N increases.[55]

In the above, it has been assumed that all resistors have equal values. However, suppose that the output shunt resistor has a value R_p instead of R. Then if X inputs are stimulated, the output becomes :-

$$V_o = V \frac{X}{R/R_p + N - 1} \qquad (2)$$

It follows that the threshold circuit is now required to differentiate between voltages differing by :-

$$\frac{V}{R/R_p + N - 1} \qquad (3)$$

The curves in fig.17.5 show the variation of this fractional voltage threshold for various values of N.

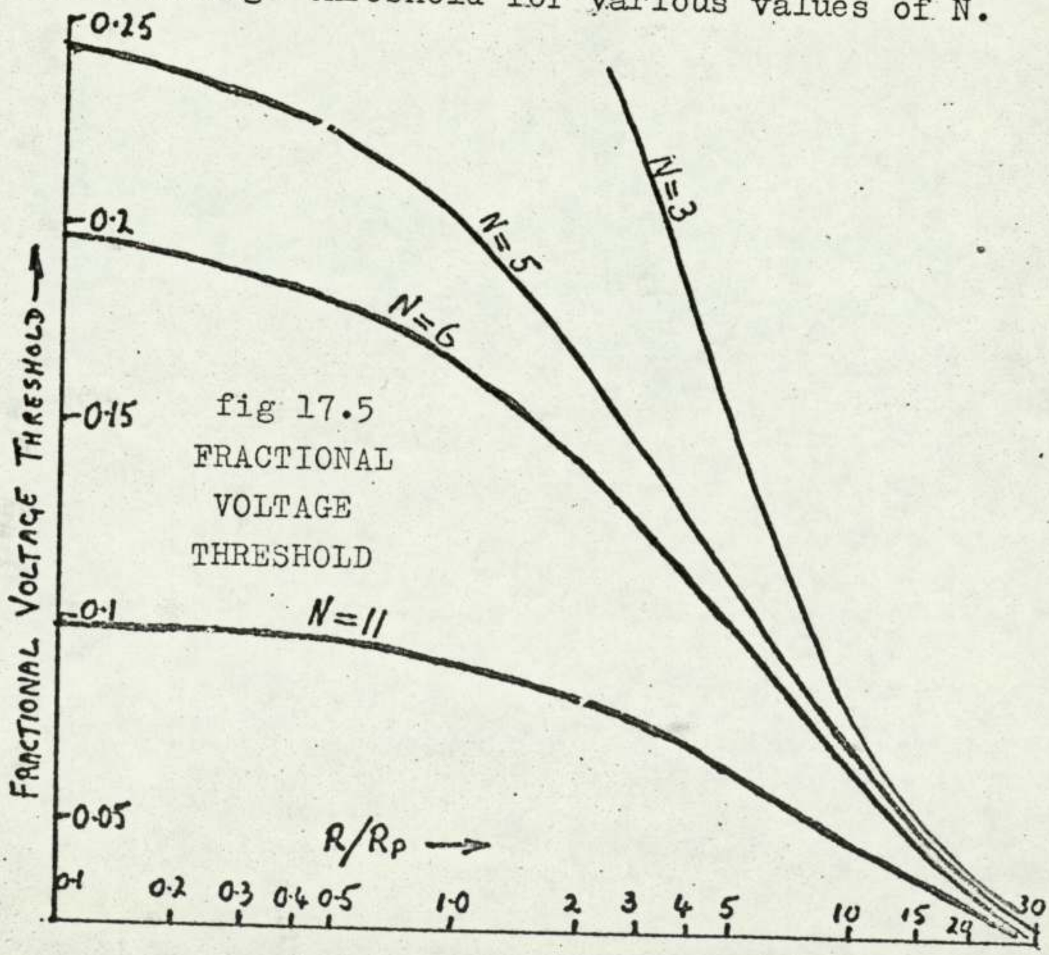


fig 17.5
FRACTIONAL
VOLTAGE
THRESHOLD

From fig. 17.5, it is seen to be advantageous if $R_p > R$. However there is not much gain beyond the value $R_p = 2R$.

Suppose now that the circuit is changed so that in the absence of an input excitation, an input resistor is open-circuited instead of being connected down to the zero line. Then with equal-valued input resistors the outputs become :-

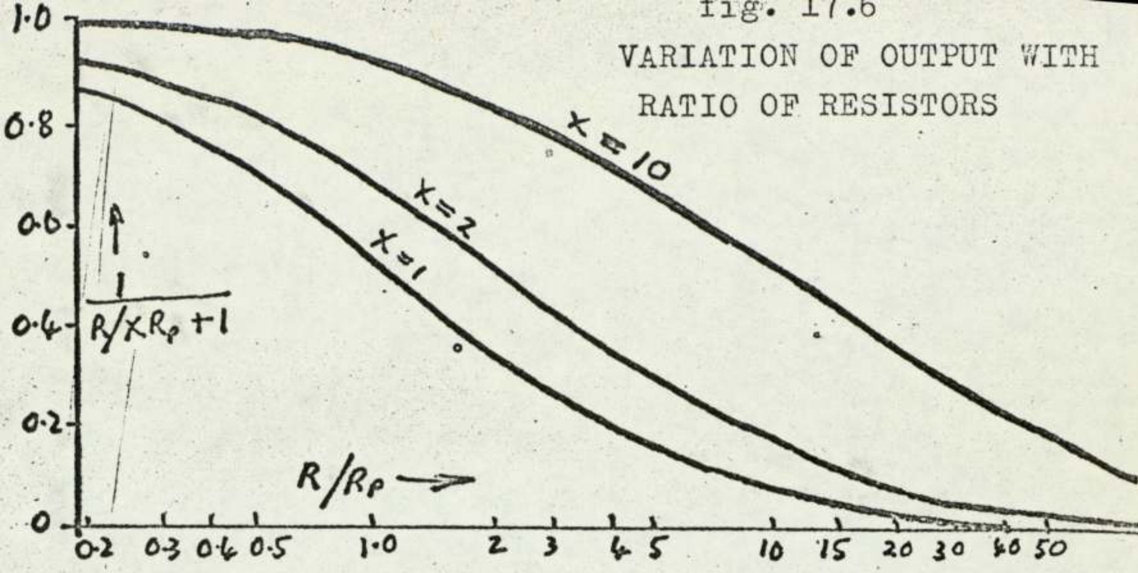
Number of inputs stimulated	Output Voltage
0	0
1	Vx1/2
2	Vx2/3
3	Vx3/4
4	Vx4/5
.....	....
N-1	Vx(1 - 1/N)

If the output resistor has a value R_p and the number of inputs stimulated is X, then the output is :-

$$V_o = V \frac{1}{R/XR_p + 1} \qquad (4)$$

The way in which this function varies with the resistor ratio R/R_p is shown in fig.17.6 for various values of X.

fig. 17.6
VARIATION OF OUTPUT WITH
RATIO OF RESISTORS



If now X+1 inputs are stimulated, then the output becomes :-

$$V_o = V \frac{1}{R/(X+1)R_p + 1} \qquad (5)$$

Consequently the threshold difference to be detected is :-

$$\text{Threshold difference} = V \frac{R/R_p}{(R/R_p+X)(R/R_p+X+1)} \qquad (6)$$

This threshold difference function is plotted against the value of R/R_p for various values of X in fig. 17.7.

It should be noted that the threshold level depends in value on the number X of inputs to be stimulated, but not necessarily on the total number (N-1) of inputs connected.

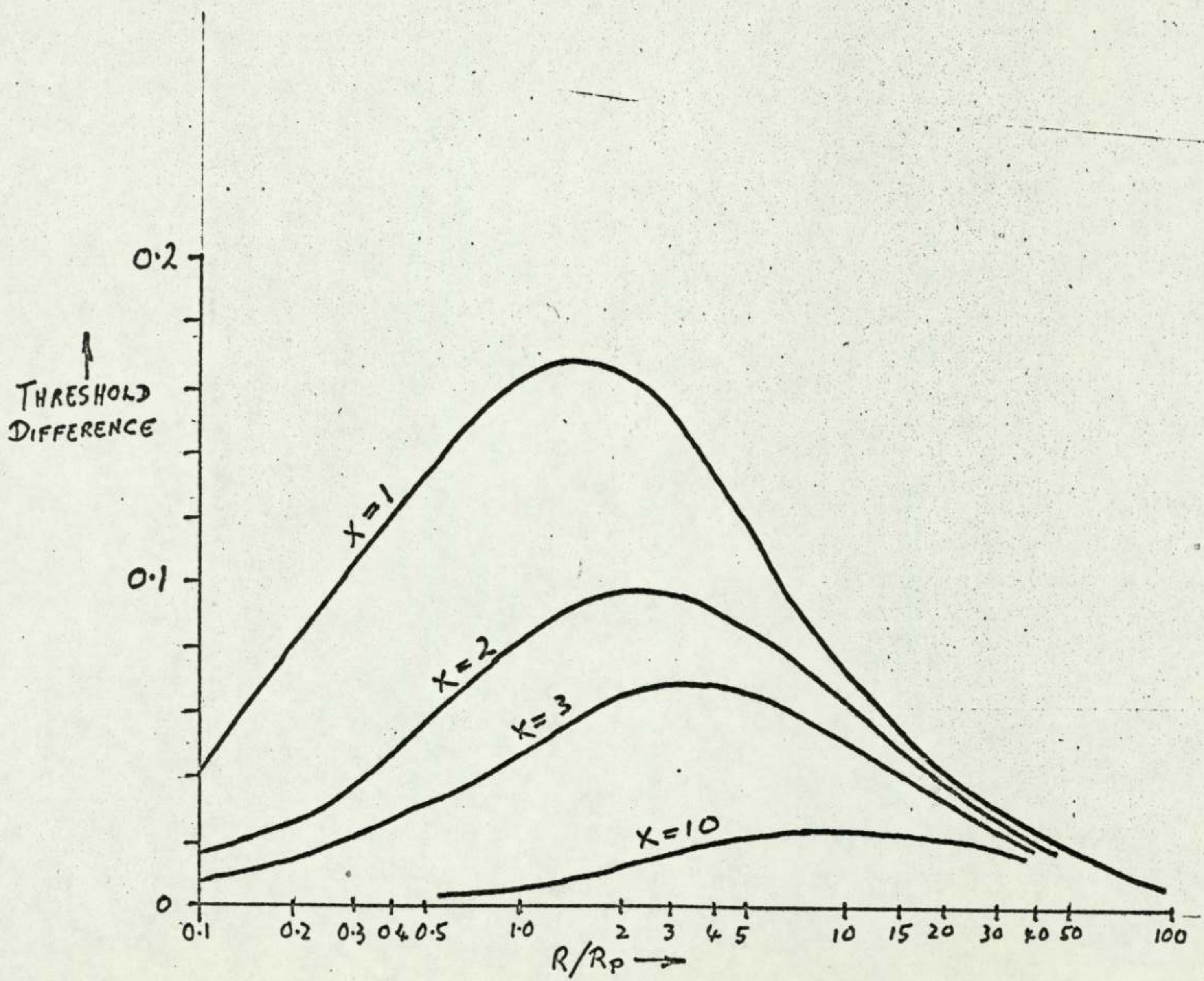


fig. 17.7 VARIATION OF THRESHOLD DIFFERENCE WITH R/R_P

With any of the methods described in this section, it is seen that for more than about ten inputs, the fractional voltage difference to be detected becomes rather small. Consequently, such circuits having a voltage-operated output could not be expected to give a good reliability if used with a large number of inputs.

A better performance can be obtained with the current-operated circuits to be described in the following sections.

17.4 RESISTOR-TRANSISTOR CIRCUITS

The form of logical switching circuit using a transistor with a number of base input resistors is well-known and it has been widely used. It is often known as a NOR circuit, though other designations have been used for the same circuit. The equations of operation of such circuits are well-known.

A resistor-transistor NOR circuit which uses a PNP type of transistor is referred to in the writer's terminology[25], (see section 16.3), as a "Positive output for One Negative input" circuit, or simply as a "P1N" circuit. Such a designation has been shown to eliminate the confusion which can be caused by the use of the term "NOR". It is possible to design such circuits so that they require two inputs to be energised before an output is given, and the writer has introduced the designation "P2N" for such "Two-or-More" circuits. An alternative brief designation for this type of circuit is derived by analogy with the term "NOR". It is "TOR", standing for "Two-or-More", and this designation is convenientto use in the present case.

The operation of the resistor-transistor TOR circuit is based on the property of the simple resistor arrangement shown in fig. 17.8.

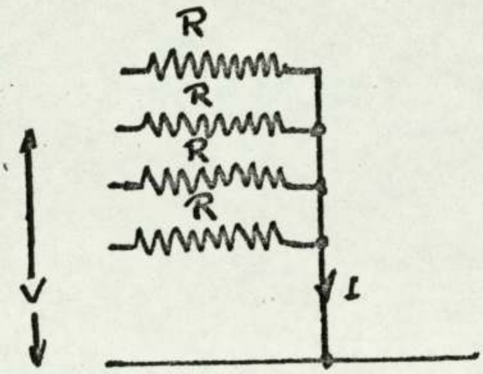


fig. 17.8 BASIC RESISTOR ARRANGEMENT

If the voltage applied to each input is zero when the input is unstimulated and is V when the input is stimulated, then with a number X of stimulated inputs, the output current is equal to XV/R. The dependence of output current on the number X of stimulated inputs is therefore linear.

The change

of output current per added output is therefore V/R and is independent of the actual number X of inputs stimulated. To construct a two-or-more circuit on this basis, it is necessary to detect a current of 2V/R or greater and to ignore currents of V/R or less. The detection threshold should therefore be set at 1.5V/R.

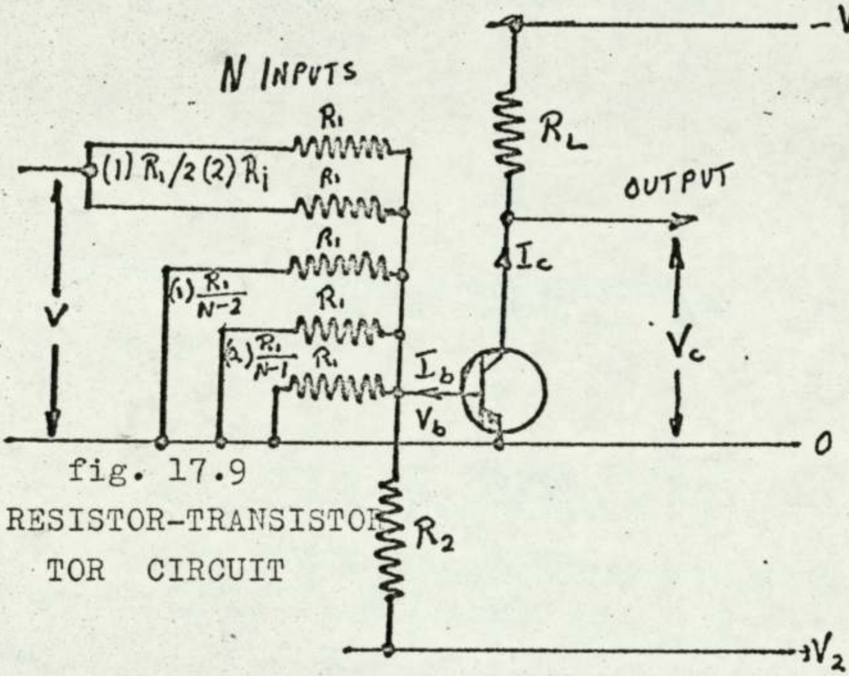


fig. 17.9
RESISTOR-TRANSISTOR TOR CIRCUIT

For the complete resistor-transistor TOR circuit of fig17.9, two conditions of operation must be fulfilled. When two inputs are simultaneously stimulated with voltage V, the transistor must carry base current I_b so that the collector conducts and the output terminal is moved positively. This requirement is expressed by equation (1), where V_{b-} is the base-to-emitter voltage of the transistor when it is carrying base current I_b (and so conducting in the collector circuit), and N is the total number of inputs.

$$\frac{V - V_{b-}}{R_1/2} = \frac{V_{b-}(N-2)}{R_1} + I_b + \frac{V_2 + V_{b-}}{R_2} \qquad (1)$$

When only one input is energised with voltage V, the transistor is required to be cut off with its base positive to its emitter by voltage V_{b+}. Its collector then only carries the low reverse leakage current I_{co} :-

$$\frac{V + V_{b+}}{R_1} + \frac{V_{b+}(N-1)}{R_1} + I_{co} = \frac{V - V_{b+}}{R_2} \qquad (2)$$

From these equations, the required value of R_1 is

$$R_1 = \frac{(V_2 - V_{b+})(2V - NV_{b-}) - (V_2 + V_{b-})(V + NV_{b+})}{I_{co}(V_2 + V_{b-}) + I_b(V_2 - V_{b+})} \qquad (3)$$

As an approximation, if $V_{b+} \ll V_2$, $V_{b-} \ll V_2$ and I_{co} is small, then :-

$$R_1 \doteq \frac{V - N(V_{b-} + V_{b+})}{I_b} \qquad (4)$$

Hence it is necessary to have $V > N(V_{b-} + V_{b+})$ (5)

and since in practice $V_{b-} + V_{b+} \ngtr$ 1/4 volts,

then the basic design requirement is $V > N/4$.

This gives a remarkably simple approach to the design of such circuits. In practice, the value of V is limited and this fact places a limitation on the maximum number of inputs which can be used.

The relationship between the values of the two resistors R_1 and R_2 is found from equation (2) :-

$$R_2 = R_1 x \frac{V - V_{b+}}{V + NV_{b+}} \qquad (6)$$

Hence as an approximation :-

$$R_1 \approx R_2 \left[1 + N\frac{V_{b+}}{V} \right] \qquad (7)$$

It has been seen above that the value of V must be related to the value of N, so write V = KN, where K is a constant. The approximation then becomes :-

$$R_1 \approx R_2 \left[1 + \frac{V_{b+}}{K} \right] \qquad (8)$$

Equations (1) to (8) give a guide to the practical design of TOR cicuits using transistors and resistors, though it is difficult to fix the values of V_{b+} and V_{b-}.

The minimum permissible value of collector load resistor R_L is determined by two factors. First, the base current must be sufficient to bottom the transistor. Secondly it is desirable to limit the maximum possible collector dissipation W_m of the transistor. Now the collector voltage is given by :-

$$V_c = V_1 - I_c R_L \qquad (9)$$

Whence, by differentiation, $W_m = V_1^2/4R_L$. The design requirement is therefore

$$R_L \geqslant \frac{V_1^2}{4W_m} \qquad (10)$$

17.5 INTEGRATED-AMPLIFIER TOR CIRCUIT of section 17.4

Equation (5) shows that the number of inputs which can be used with a resistor-transistor TOR circuit is less than

$$N < \frac{V}{V_{b-} + V_{b+}} \qquad (1)$$

This can be written as :-

$$N < \frac{V}{V_p} \qquad (2)$$

where V_p is the smallest range over which the input base-to-emitter voltage of the transistor must swing in order to change the state of the transistor from the fully-on to the fully-off condition. Consequently, the smaller the required base-to-emitter input voltage swing required, the greater the number N of inputs which can be used. Now the required voltage swing is a function of the voltage gain of the transistor used.

Integrated amplifiers having a high voltage gain have recently become available at low cost. The possibility therefore occurs of using one of these to replace the amplifying transistor in a TOR circuit. The possibility was tried out using an integrated amplifier type SL701c.

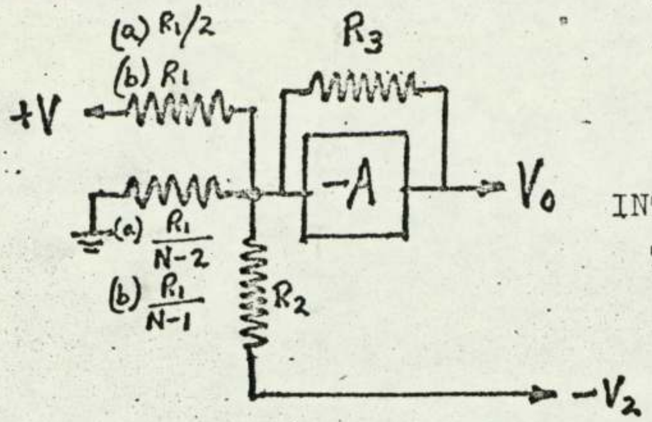


fig. 17.10
INTEGRATED-AMPLIFIER
TOR CIRCUIT

The basic circuit is as shown in fig. 17.10.

If it is assumed that the range of amplifier input voltage swing required is very much less than either V or the maximum positive or negative values of V_o, then analysis of the arrangement shows that:-

$$\frac{2V_o}{R_3} = \frac{V - V_b(2N-3)}{R_1} \qquad (3)$$

where N is the number of inputs and V_b is approximately one half of the required input voltage swing. Also, $V_o \approx V$. It follows that it is necessary to have $V > V_b(2N-3)$. (4) Now $V/V_b \approx A$, and if $N \gg 1$ then $N < A/2$. This means that, since A is large, then majority logic circuits having very many inputs can be constructed.

Using the SL701c amplifier shown in fig. 17.11, it was quite possible to handle 100 inputs and to obtain reliable two-or-more operation. No attempt was made to obtain the ultimate performance in these tests, since it was never necessary to work with more than twenty inputs in the initial work. Moreover, the design of integrated amplifiers is progressing continuously at the present time, while the cost is falling rapidly.

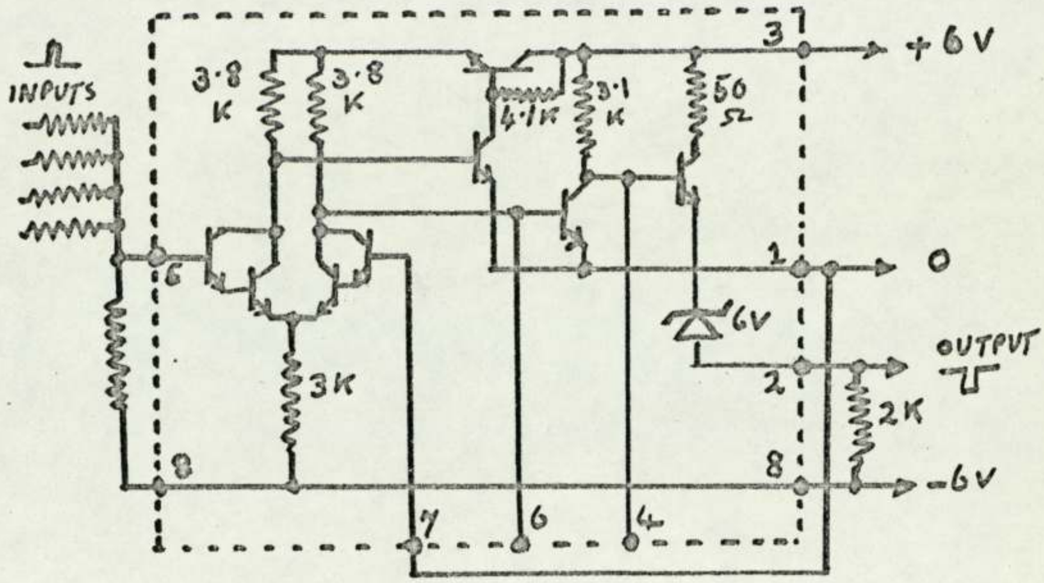


fig. 17.11 SL701c INTEGRATED AMPLIFIER IN TOR CIRCUIT

In the writer's terminology,[25] the circuit of fig. 17.11 is called an N2P circuit. With arrangements such as this, the writer's students have carried out preliminary work on pulse-height standardisation by use of a Zener diode in each input circuit of a TOR arrangement.

17.6 INHIBITORY INPUTS TO TOR CIRCUITS

With the TOR arrangement described earlier, all input voltages are equal to either 0 or +V. A negative output is obtained when there are two-or-more positive inputs. It is possible to make some of the inputs inhibitory by connecting them to an input voltage of -V. We simply put $-V_2 = -V$, and $R_2 = R_1/M$ in fig. 17.10 , while we have L inputs connected to voltage +V and N-M-L inputs connected to ground. In these circumstances, if the gain A is large so that V_b is small

then if L = M-1, the output voltage is $V_o = VR_3/R_1$;

if L = M, the output voltage is $V_o = 0$;

and if L = M+1, the output voltage is $V_o = -VR_3/R_1$.

It is convenient to set an output threshold voltage of $-VR_3/2R_1$, an output being indicated if the voltage is more negative than this value. In this case, an output is obtained if $L \geqslant M+1$. If only a single resistor $R_2 = R_1$ is fitted as in the earlier case, then M = 1 and the arrangement becomes a two-or-more circuit which gives a negative output voltage if ever the number of inputs L taken to a positive supply exceeds two.

17.7 TOR CIRCUIT COMPONENT AND VOLTAGE TOLERANCES

It is here convenient to consider the input conductances rather than the resistances. In the TOR input circuit shown in fig. 17.12, suppose that the fractional tolerance on each conductance is g, so that the value of each conductance lies in the range $(1 \pm g)G$. Also suppose that the actual tolerance on each input voltage is v, so that the value of each input voltage lies in the range $(V \pm v)$. Also let the voltage tolerance on each input which is nominally grounded also be v, so that a "ground" point is actually at a voltage in the range $\pm v$.

Then the maximum input current with only one input is :-

$$I_{max} = (V+v)(1+g)G + (N-1)v(1+g)G - I_k \qquad (1)$$

while the minimum input current with two inputs is :-

$$I_{min} = 2(V-v)(1-g)G - (N-2)v(1+g)G - I_k \qquad (2)$$

Here N is the total number of inputs and I_k is the constant bias current. It is necessary to ensure that

$$I_{min} > I_{max} \qquad (3)$$

From these relationships, the maximum number of inputs is given by :-

$$N < \frac{V/v + g}{2(1 + g)} \approx \frac{V}{2v} \text{, since g is small.} \qquad (4)$$

Thus under the conditions stated, the maximum number of inputs is determined mainly by the voltage tolerance.

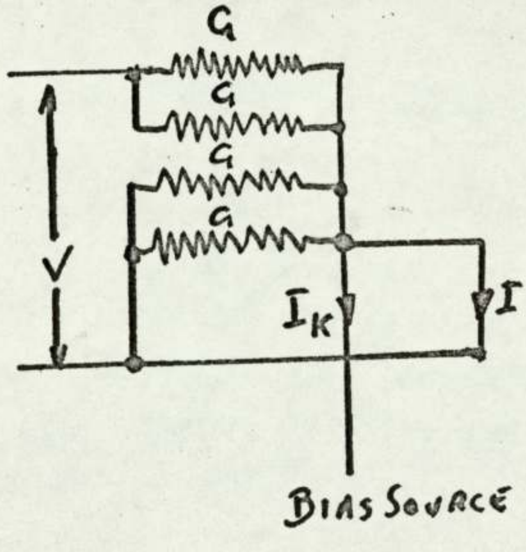


fig. 17.12

TOR CIRCUIT FOR CALCULATION
OF MAXIMUM NUMBER OF INPUTS.

17.8 USE OF FIELD-EFFECT TRANSISTORS

The use of field-effect transistors[46] as the input switches in two-or-more threshold gates was investigated briefly.

Two forms of circuit were tried in the course of the present work. One circuit used N-channel F.E.T.'s while the other used P-channel F.E.T.'s. The circuit diagrams of these arrangements are given in fig. 17.13 and in fig. 17.14. The circuits are required to give a negative-going output pulse if two-or-more inputs receive negative pulses simultaneously.

The first F.E.T. circuit, shown in fig. 17.13, incorporated a number of N-channel devices, all normally in the conducting state. An input negative pulse to two or more of the F.E.T. gates gives at the commoned drain connection a positive pulse of sufficient amplitude to produce a negative output pulse. This arrangement was used for a time as part of the main machine, but the characteristics of the transistors did not prove to be stable enough for normal use and eventually the arrangement was abandoned.

The other F.E.T. circuit, shown in fig. 17.14, used P-channel devices which conduct only when a negative pulse is applied to their gate electrodes.

Because of a shortage of P-channel F.E.T.'s at the time, a complete circuit was not built.

It was found to be difficult to construct stable and reliable circuits because of troubles with the low-cost field-effect devices used. It was consequently decided not to use these forms of gate for initial work on the system. These circuits are felt to be worthy of further investigation, and it is hoped to continue this work on field-effect threshold gates at a later date. Such arrangements would probably be fairly easy to adapt to integrated circuit form.

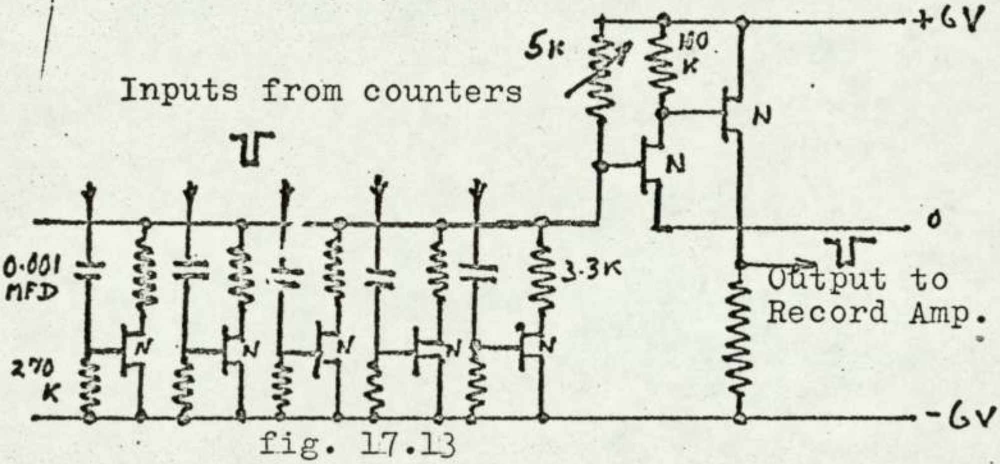


fig. 17.13

Two-or-More circuit using N-channel FET's

All FET's are type 2N3819.

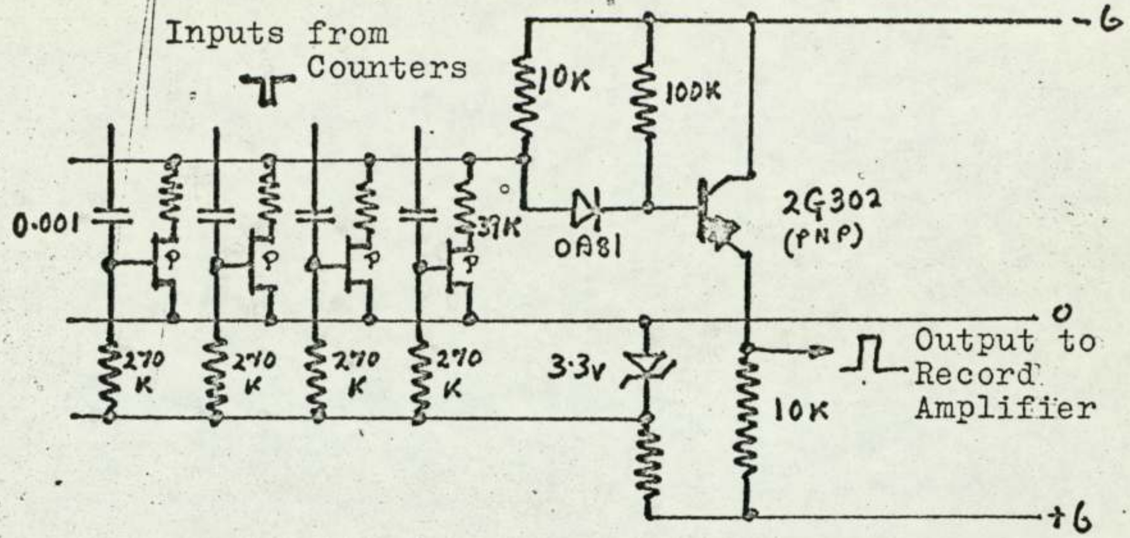


Fig. 17.14 Two-or-more circuit with P-channel FET's.

All FET's are type 2N3820.

17.9 DIODE TWO-OR-MORE CIRCUIT

It is possible to use diode AND and OR gates in a two-or-more majority logic circuit. For example with four inputs A,B,C,D, the required operation is :-

$$A.B + A.C + A.D + B.C + B.D + C.D$$

To carry out this operation with N inputs, a total of $\frac{N(N-1)}{2}$ AND gates is required. Each AND gate must have two input diodes, and each is followed by a diode which forms part of an output OR gate. Consequently, to implement this approach, a total of $\frac{3N(N-1)}{2}$ diodes is required. In fig. 17.15 this is plotted against N. It is seen that the number of diodes required increases very rapidly as the number of inputs is increased.

While this approach is consequently not practicable for use with a large number of inputs, it was convenient for use in initial work. The diodes were plugged into a matrix board to form the AND gates. The output from the OR gate was amplified and phase-inverted, and the resulting output coincidence pulses were shaped by an integrated monostable circuit before being passed to the record amplifier. The circuit used is shown in fig. 17.16.

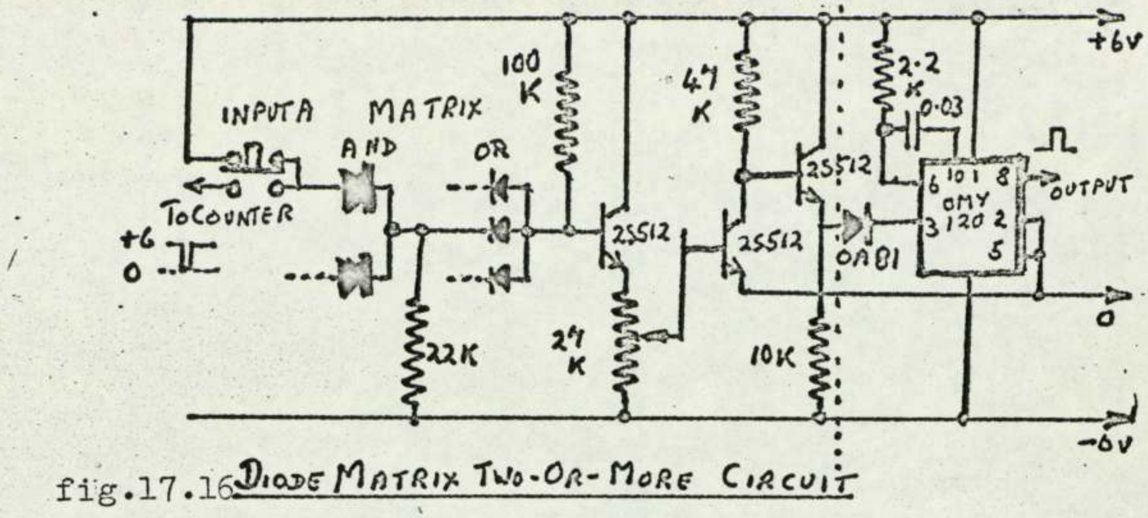


fig.17.16 Diode Matrix Two-Or-More Circuit

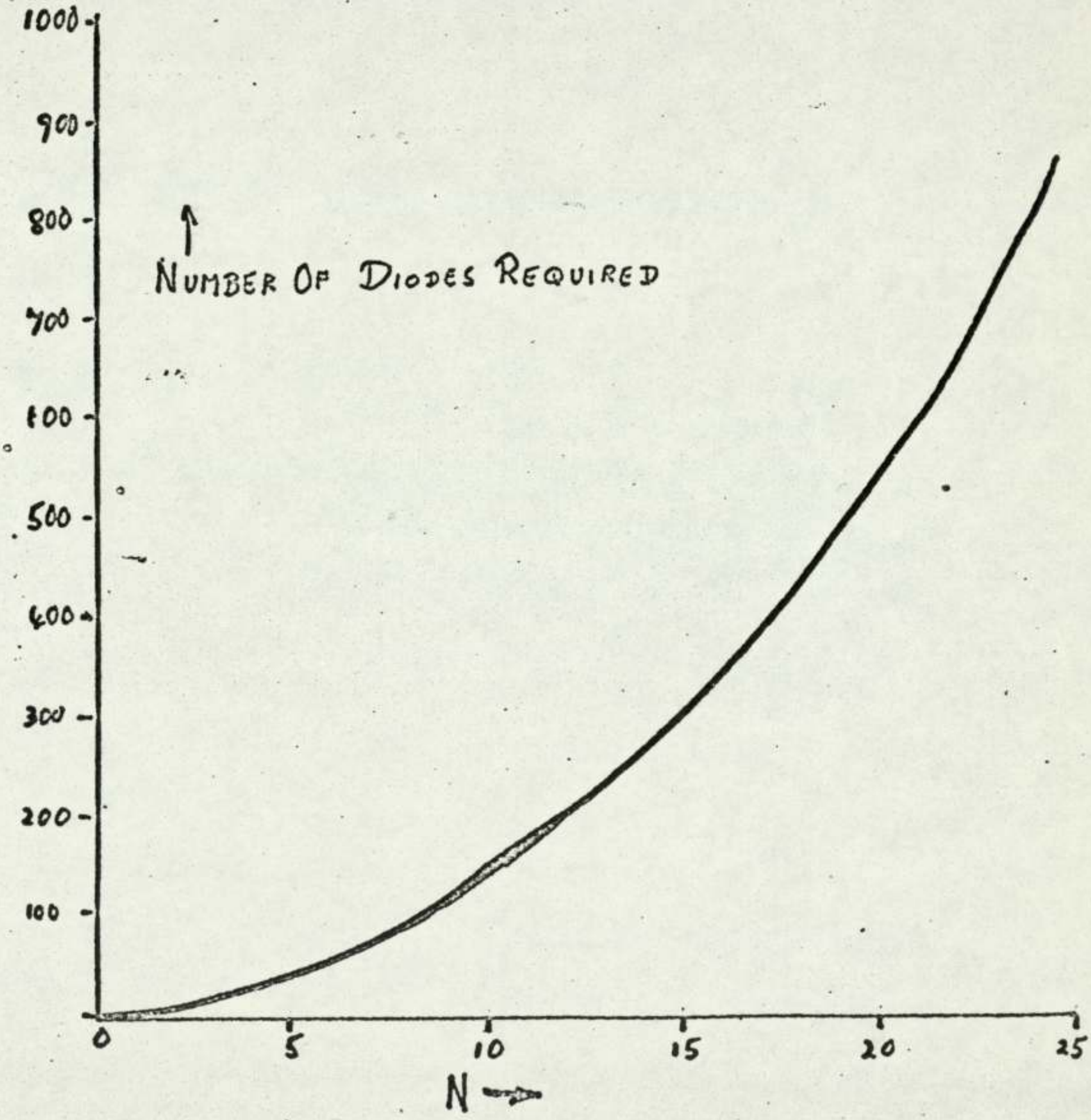


fig. 17.15. Number of Diodes for Two-or-More Circuit.

18. OUTPUT CIRCUITS

18.1 OUTPUT CIRCUITS : INITIAL WORK

The logical requirement for an output signal a_o to be required from one particular output terminal A_o of the machine is:"an output is required if ever the corresponding input circuit A_1 is stimulated OR if another input circuit B_1 is stimulated AND there is a signal $(a_1b_1)_m$ from the memory circuit indicating past association of input signals a_1 and b_1". This logical requirement can be stated concisely as :-

$$a_o = a_1 + b_1(a_1b_1)_m \qquad (1)$$

This requirement is not difficult to implement with logical circuits. However the resulting output is then in the form of very short pulses. These must be converted to longer pulses which can be used to switch on output devices such as indicator lamps.

In the initial work, a three-input AND gate was fitted to the input of each of a number of monostable pulse-length converters. Each AND gate carried out the logical operation :-

$$a_x = b_1(a_1b_1)_m \qquad (2)$$

This operation would only require a two-input AND gate, but it would not be selective. In order to ensure that the pulses are routed only to one specific output A_o, it is necessary to add a further set of input pulses a to a third input circuit A of each AND gate. Then :-

$a_o = a.a_x = a.b_1(a_1b_1)_m.(3)$. Fig.18.1 shows the original output monostable circuit. A positive signal AND gate operates directly into the base circuit of a transistor which is used to switch the output lamps for indication. When all three inputs are simultaneously positive, the first transistor conducts and its collector moves negatively. Since the capacitor is initially charged up almost to the full supply voltage, the base of the second transistor is at this time taken negative to its emitter. Consequently, the second transistor cuts off and its collector moves positively. This positive change is coupled back to the negative end of the base bias chain of the first transistor. Consequently the first transistor continues to conduct until the coupling capacitor C has discharged (via the base bias resistor of the second transistor) sufficiently for the second transistor to conduct again. The first transistor is then restored to the non-conducting state and the output indicator lamp is extinguished.

Once triggered, the output therefore persists for a time determined by the coupling time constant CR. The output indicator lamp is taken to a 24 volt supply while the rest of the circuit is supplied from 6 volts. This arrangement is intended to limit mutual interference between the various monostable circuits and to enable reasonably low-current 24 volt indicator lamps to be used.

The monostable circuits had to be reasonably simple and inexpensive since large numbers of them are likely to be required. The arrangement described above was used in early work. Later changes in the rest of the equipment made it desirable to operate the input AND gate on coincidence of negative rather than positive pulses. Consequently the monostable circuit was modified, with the input gate now taken to the second transistor rather than to the load-switching transistor, as shown in fig. 18.2 .

At the same time, the organisation of this part of the system was changed. A central AND gate now detects coincidence between any stimulated prime-number input and any past-association product signal which is coming from the memory circuits. Any resulting output pulses from this central AND gate are passed to all of the AND gates of the output monostable circuits. The AND gates feeding the output monostable circuits are now only required to have two inputs. One input of each AND gate is supplied with the pulse signals from the central AND gate. The other input is supplied in each case with pulses from the appropriate prime-number counter.

Thus each monostable circuit is operated by pulses from an AND gate which acts as a selector.

In order to obtain a reasonably long monostable pulse for application to the indicator lamp, quite a large value of timing capacitor is required in the monostable circuit. When the output-switching transistor turns off in the circuit of fig.18.1, the coupling capacitor recharges via the low-resistance load and the base of the second transistor. The resulting heavy base current is undesirable, and it is prevented from flowing in the new circuit by the addition of a diode in the base- coupling circuit. Because of this there is a fairly long refractory time while the capacitor is recharging before the circuit can again give the same length of output pulse.

In the experimental arrangement, the additional a_1 term in the required logical statement is provided by spare contacts on the input push-buttons. It would not be difficult to replace these contacts with semi-conductor switches, but it was not worth-while in the initial experimental work. The push-buttons used are of the illuminated type, and the push-button lamps are connected in parallel with the main indicator lamps.

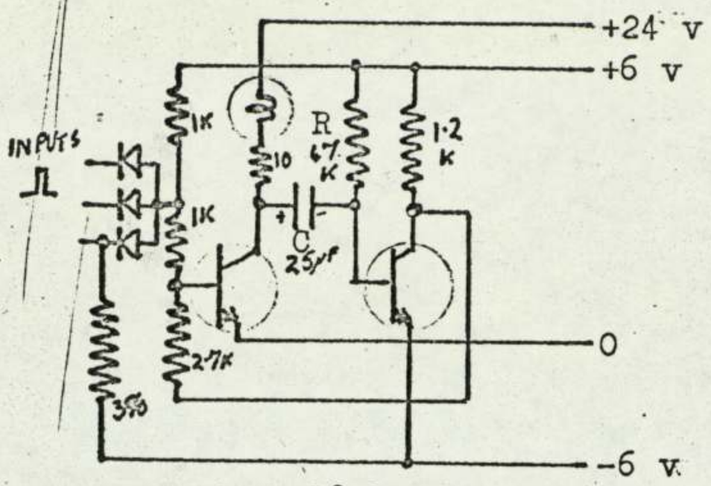


fig. 18.1

18.1 First Output Monostable Circuit.

Transistors are type BFY 51

Diodes are type OA81

This circuit was intended to work with positive inputs

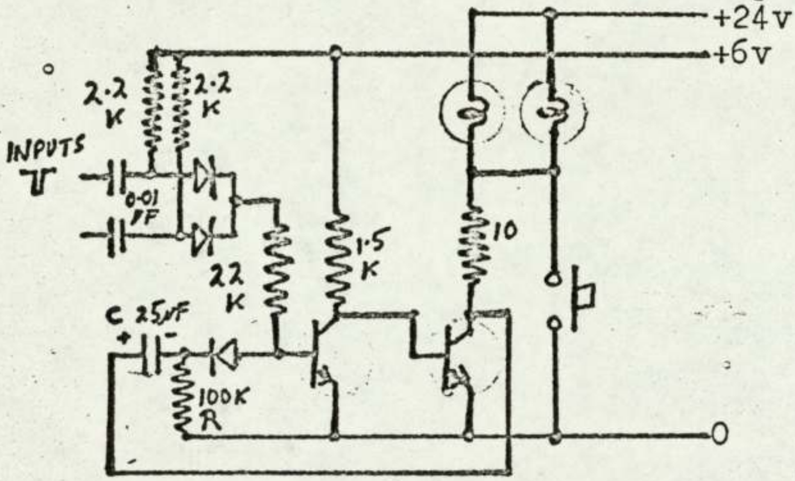


fig. 18.2

Second Output Monostable Circuit

Transistors are type BFY51

Diodes are type OA81

This circuit works with negative inputs.

18.2 FINAL OUTPUT CIRCUIT

The output circuits described in section 18.1 were designed in an attempt to achieve economy. However, occasional problems of interaction between units were still encountered. It was therefore decided to separate the function of lamp switching from that of monostability.

An integrated circuit is now used as the monostable device, as shown in fig. 18.3. The input is coupled to the input AND gate by an emitter-follower. The output of the monostable circuit is isolated from the base circuit of the lamp-switching transistor by another emitter-follower. A tendency to spontaneous continuous oscillation was encountered in some of the integrated circuits used. Since a rapid action is not required in these monostable circuits, it was possible to ensure the avoidance of oscillation by the addition of a capacitor to the negative line.

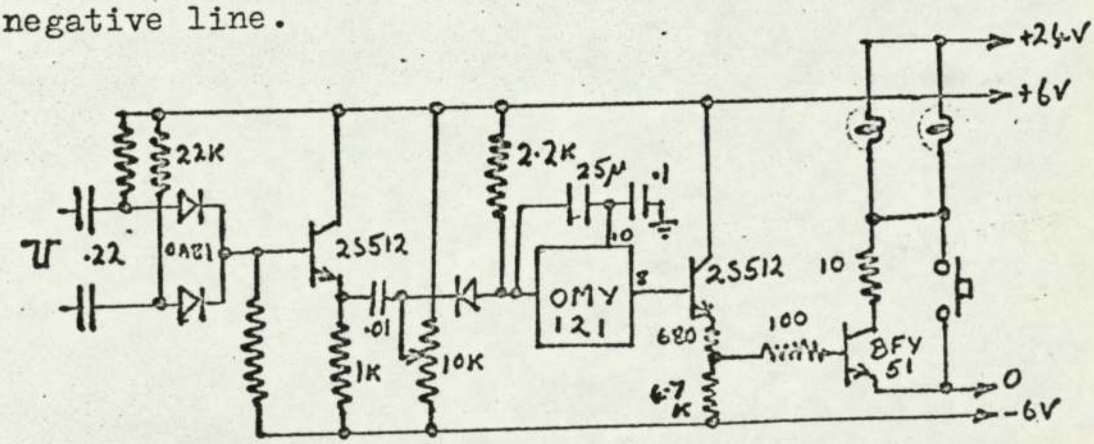


fig. 18.3. FINAL OUTPUT CIRCUIT

19. CENTRAL MEMORY "AND" GATE

19.1 CENTRAL MEMORY "AND" GATE.

As stated in section 18.1 , a central AND gate is now used to detect coincidence between any stimulated prime-number input and any past association product signal which is coming from the memory circuits. When any one input is stimulated, pulses come from the appropriate counter via the input diode OR gate to a diode AND gate, as in fig.19.1. A double emitter-follower then feeds a monostable pulse-shaping circuit. The shaped output pulses are taken, via a power-amplifying and phase-inverting transistor, to the output units.

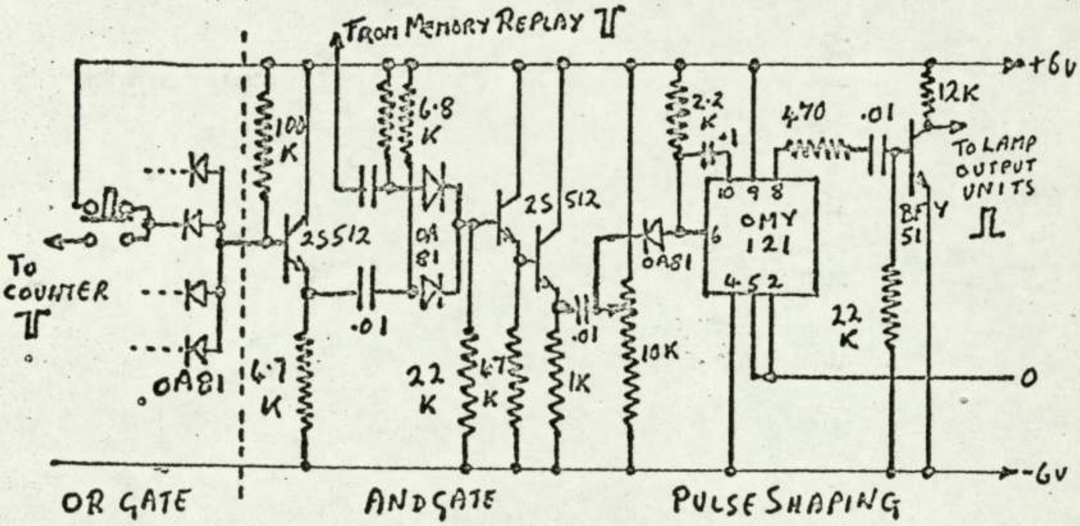


fig. 19.1 CENTRAL MEMORY AND GATE

20. MAGNETIC RECORDING

20.1 MAGNETIC PULSE RECORDING

In the course of the present work, it was necessary to investigate methods of recording short pulses on magnetic tape or on a magnetic drum. Those methods which have been used in digital computer storage systems are not automatically applicable.

In the prime-period pulse system described in section 15, it is required that new information can easily be added to the information impressed earlier on the recording medium. Consequently, methods based on the RZ or Return-to-Zero method of recording are desirable.[47] Unfortunately, the minimum permissible spacing between adjacent recorded signals is greater with this system than with some other systems. This fact is often expressed by the statement that the maximum bit density of this system is low.

In an attempt to optimise the storage, it was necessary first to consider the effect of bit-density variation in the particular form of magnetic recording likely to be adopted in the present association memory system.

In this, the pulses of information to be stored are applied to the record head via a capacitor from an amplifier supplying high-voltage record pulses.

If the voltage applied via a capacitor to the recording head has a rectangular waveform, then the current flows in two opposing quasi-differential spikes. If these current spikes have sufficient amplitude to take the recording medium into saturation, then the magnetisation of the recording medium takes the form of two opposing pulses which can be considered as a first approximation to be rectangular. When a tape having this magnetisation moves past a replay head, the resulting output voltage takes the form of two opposing bipolar pulses.

If the applied recording pulse is progressively narrowed, the two output voltage pulses approach until they combine into a single larger amplitude negative pulse. It is then of interest to consider whether there is any optimum width of recording pulse of voltage.

It is usual to assume that the read-head voltage response produced by a step change of tape magnetisation can be approximated by a Gaussian function :-

$$f(t) = \exp(-t^2) \qquad (1)$$

This proves in practice to be a reasonable approximation. Now if we have two adjacent step-changes in opposite directions, the read pulse is approximated by :- $f(t) = \exp(-t^2) - \exp(-(t-x)^2)$, (2)

where x is the time between the step changes.

If we have four adjacent step changes in +, -, -, +.

directions, the read pulses are approximated by :-

$$f(t) = \exp(-t^2) - \exp(-(t-x)^2) - \exp(-(t-T)^2) + \exp(-(t-x-T)^2) \qquad (3)$$

In the case under consideration, x is the width of the magnetising pulses, while T is the width of the applied voltage pulse.

It is required to find the optimum value of T, with x fixed, to give a maximum amplitude of read pulse. Consider the amplitude at the mid-point where $t = \frac{T+x}{2}$. Substitution of this value into equation (3) shows that the maximum amplitude is given by :-

$$f_{max}(t) = -4 \exp\left(-\left(\frac{T^2+x^2}{4}\right)\right) x \sinh \frac{Tx}{2} \qquad (4)$$

It can be assumed that x is fixed at the shortest possible value. Now put T = kx , then :-

$$f_{max}(t) = -4 \exp\left(-\left(\frac{k^2+1}{4}\right)\right) x \sinh \frac{kx^2}{2} \qquad (5)$$

In fig.20.1, $f_{max}(t)$ is plotted against the value of k It is seen that there is a maximum value at k=1.5. The maximum value is given by :-

$$f_{max}(t) = 2(\exp(-1.56x^2) - \exp(-0.56x^2)). \qquad (6)$$

If the value of x is taken as unity, then evaluation of equation (6) gives :-

$$f_{max}(t) = -1.46 \qquad (7)$$

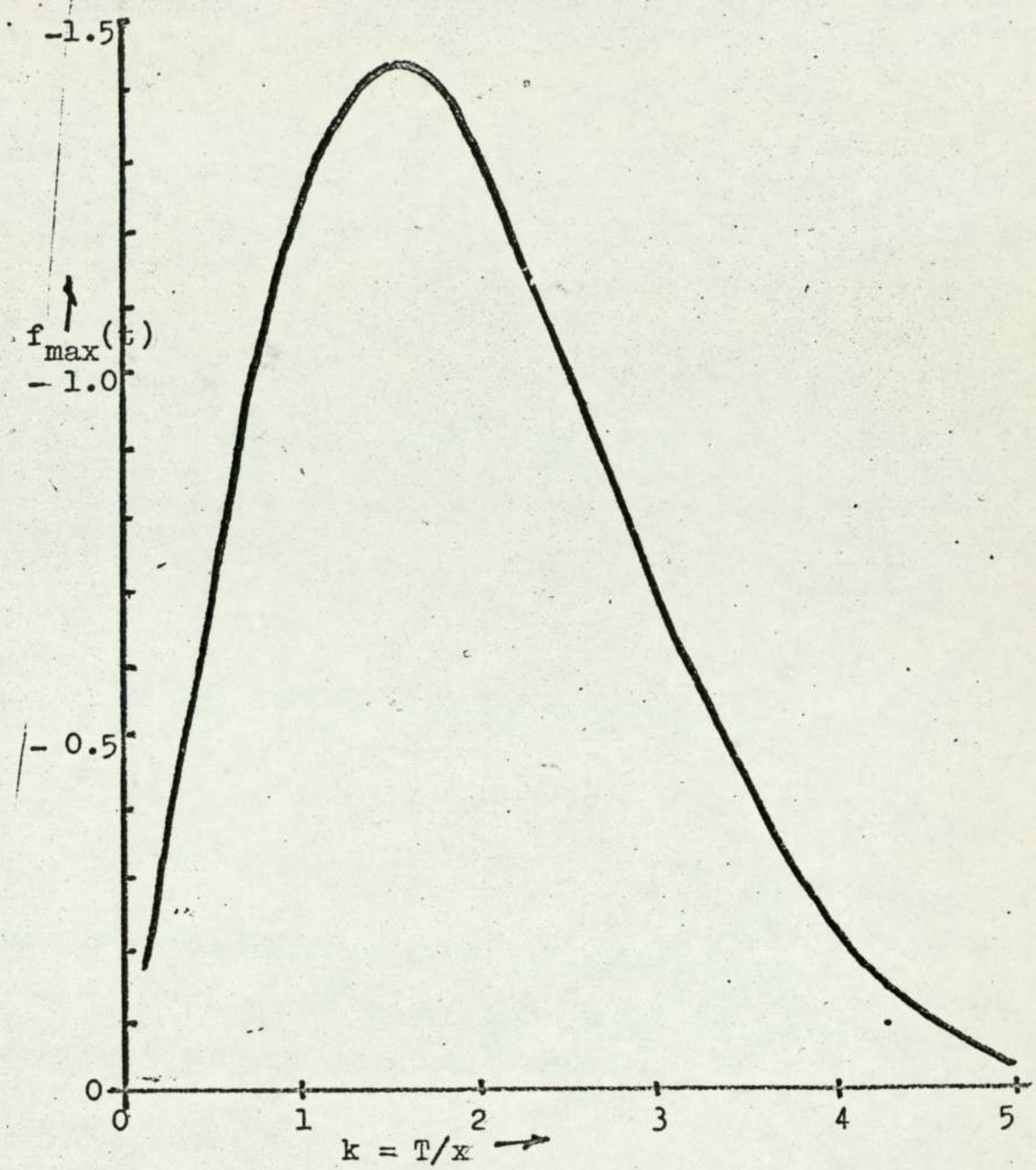


fig. 20.1 VARIATION OF $f_{max}(t)$ WITH k

It has been shown above that the optimum pulse width is given by T = 1.5 x , if maximum peak amplitude is taken as the criterion. It is of interest to examine the way in which the actual shape of the read pulse changes as the value of T/x is varied. Curves showing the pulse shape for various values of T/x, as calculated from equation (3), are given in figures 20.2 and 20.3.

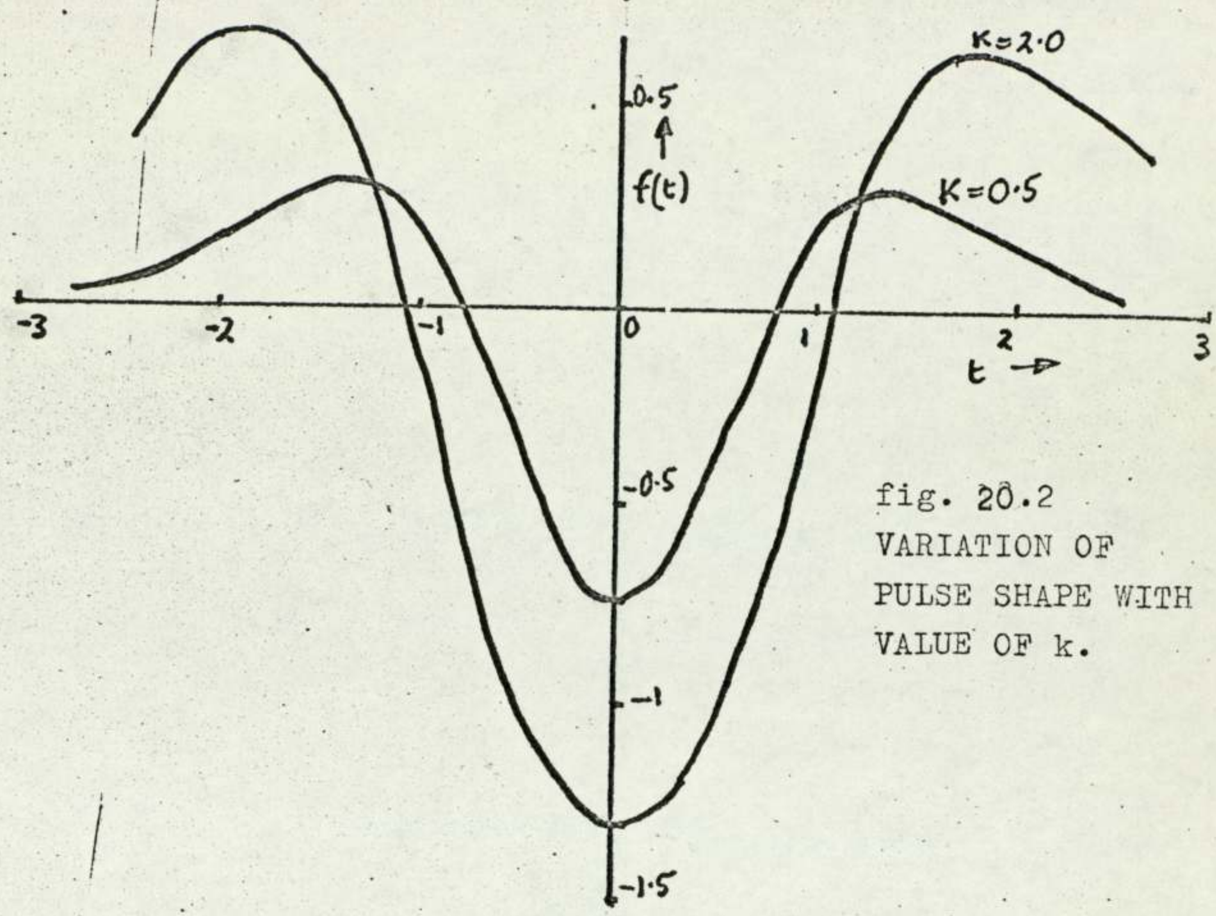


fig. 20.2
VARIATION OF
PULSE SHAPE WITH
VALUE OF k.

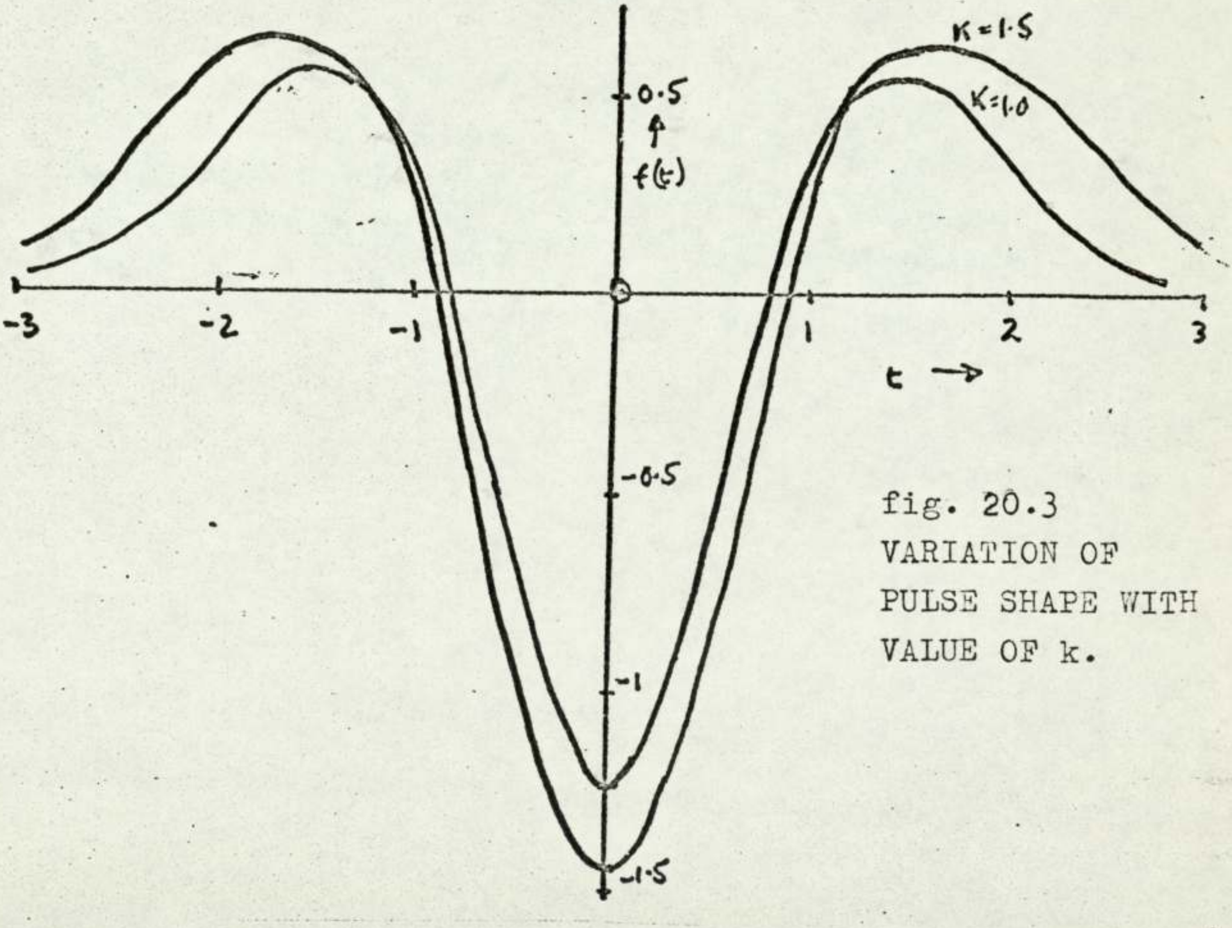


fig. 20.3
VARIATION OF
PULSE SHAPE WITH
VALUE OF k.

In practice, the positive part of the waveform can easily be eliminated, for example by diode switching. It is then of interest to examine the behaviour of the ratio :-

$$\frac{\text{Negative Peak Amplitude}}{\text{Distance Between Zeros}},$$

and to use this ratio as a figure of merit as the value of T is varied. The figure of merit varies with the value of k = T/x as shown in fig. 20.4.

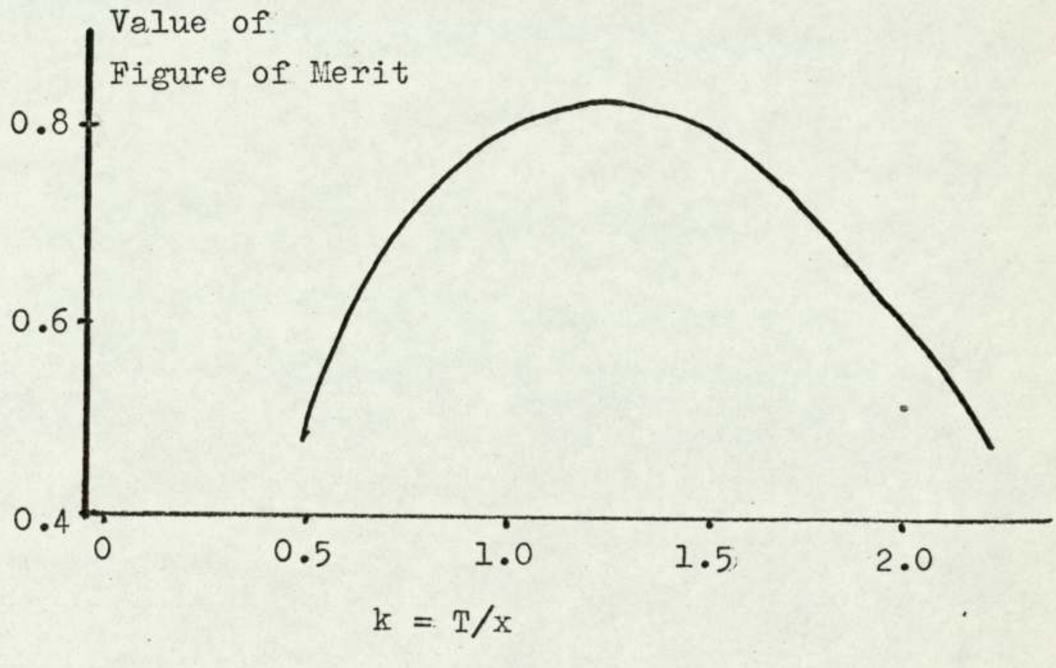


fig. 20.4 VARIATION OF FIGURE OF MERIT WITH k

The preceding analysis is of use for well-designed digital magnetic recorders having low-inductance head windings. It had been hoped that such a recorder could be obtained in the course of the present work. Unfortunately, this was not possible, and it was necessary instead to make use of a low-frequency recorder designed for audio-frequency, non-pulse applications.

Because of this, both the head inductance and the head gap lengths were excessive, and in such circumstances it is not possible to approach the optimum conditions considered above. It is hoped that the analysis can be used in future work when a properly designed digital recorder has been obtained.

In the above, it has been assumed that the basic waveshape is Gaussian. However, this provides no more than a suitable approximation to the waveshape actually obtained. An alternative approximation to the waveform is provided by the second integral of a rectangular wave. This point is likely to be of importance in future analysis of magnetic pulse recording.

20.2 RECORD AMPLIFIER

The record heads on the tape deck used have an inductance of 28 mH, and it is necessary to apply quite high voltage pulses in order to force through them a reasonable value of pulsed recording current. In early attempts, a single C407 transistor, operating with a 120 volt supply, was used to produce the pulses. However, it became clear that some pre-amplification was necessary in order to produce output pulses when the input pulses had low amplitude. In addition, the output impedance was reduced by incorporation of an emitter-follower using parallel transistors as in fig.20.5.

On the memory track, the same head is used for recording and for replaying, and so it is necessary to consider the first stage of the replay amplifier at the output from the record amplifier. A diode is added between base and emitter to protect the transistor from negative voltage spikes produced by the rate of change of current in the inductive head. A positive input pulse from the two-or-more detector produces a positive output pulse of high amplitude to the head. The inductance of the head and charge storage in the transistor causes the head current to flow for a time after the termination of the input voltage pulse. A negative voltage pulse of lower amplitude consequently follows the large positive record voltage pulse appearing across the record head.

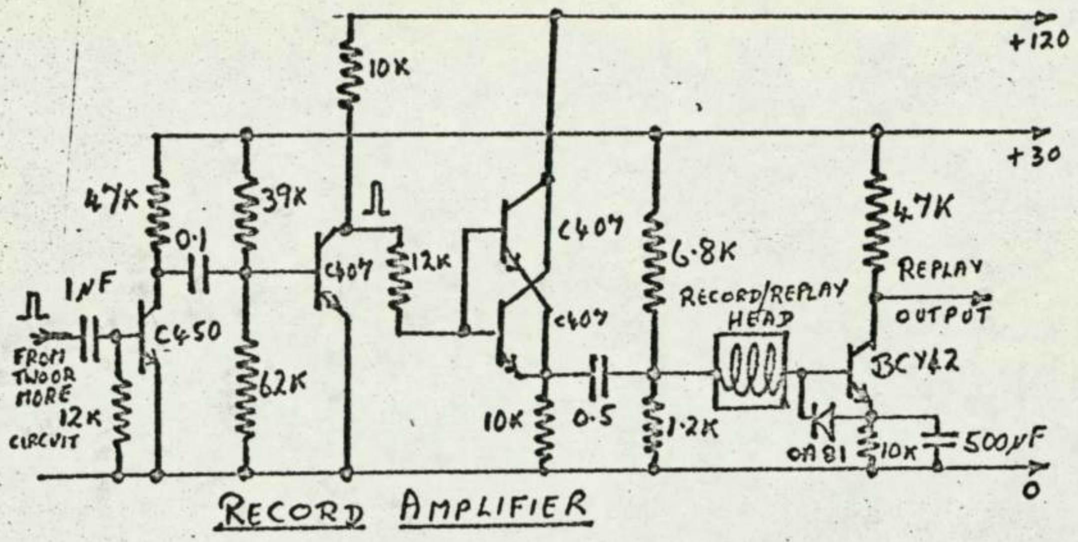
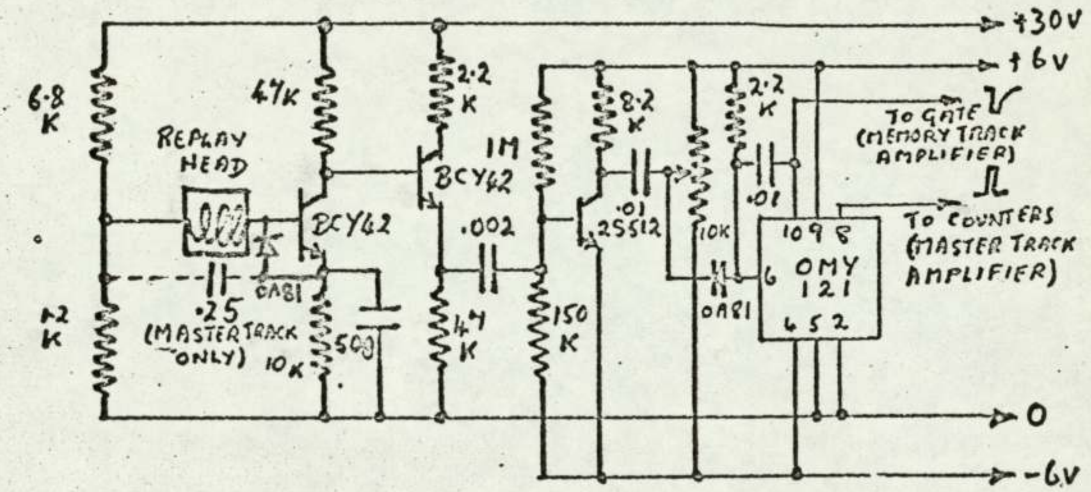

+120
+30
0
10K
47K
39K
0·1
C407
12K
C407
C407
6·8K
47K
REPLAY
OUTPUT
1NF
C450
FROM TWO OR MORE CIRCUIT
12K
62K
10K
0·5
1·2K
RECORD/REPLAY HEAD
BCY42
OA81
10K
500μF
RECORD AMPLIFIER

fig. 20.5

20.3 REPLAY ARRANGEMENTS

As shown in fig. 20.6, the master track and the memory track have almost identical replay arrangements. The first stage of the memory track replay circuit has been described in section 20.2, since it is closely linked to the record amplifier. The first stage of the master track replay amplifier uses a slightly different biasing arrangement. The output from the first stage is taken via an emitter follower to a phase-inverting amplifier which controls the input to an output pulse-shaping circuit, using an integrated monostable circuit.



REPLAY ARRANGEMENTS

fig. 20.6

20.4 END-OF-TAPE DETECTION

The magnetic tape is in the form of a loop so that any required segment from the complete memory recording is regularly available for use within a short interval. No pulses are recorded on the master track of the tape around the join in the tape loop. At this point, all counters must be reset to zero so that they all start to count in synchronism again at the start of playback of the master track.

The resetting to zero is carried out by a special end-of-tape detector circuit. In the first circuit, of fig.20.7, all the time that positive pulses are coming from the master track, the capacitor in the source circuit of an input source-follower is maintained charged to a positive voltage. At the join in the tape, the master track pulses cease for a time. The capacitor in the input source-follower circuit is then allowed to discharge via its shunt resistor. The negative-going discharge voltage triggers the monostable circuit, which produces a long negative-going pulse at its output. This pulse is amplified and inverted by a buffer amplifier and then supplied to the resest circuits of all counters.

This circuit was used successfully in all of the initial work. It was, however, rather susceptible to mal-operation caused by noise from the tape or elsewhere in the system. For the present, therefore, an electro-mechanical method of operation has been adopted. A length of metal foil is cemented to the tape at the join. Contact of a probe with the foil produces a negative signal which is used to cause the monostable circuit to give a resetting pulse as in the earlier circuit. The new circuit is shown in fig.20.8.

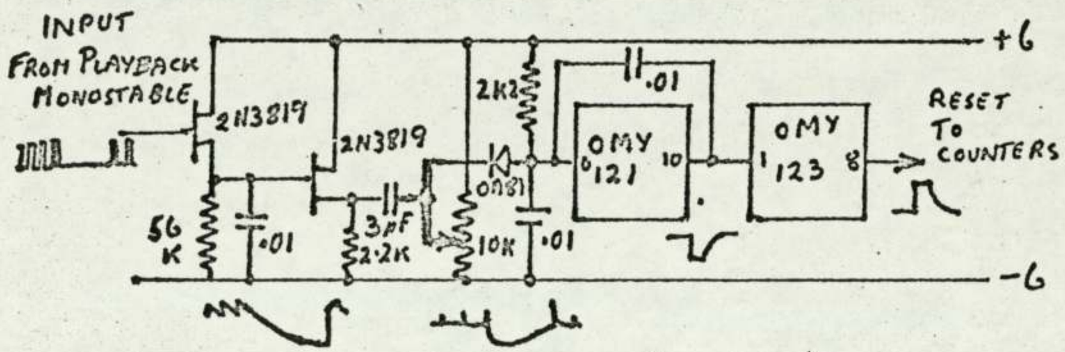
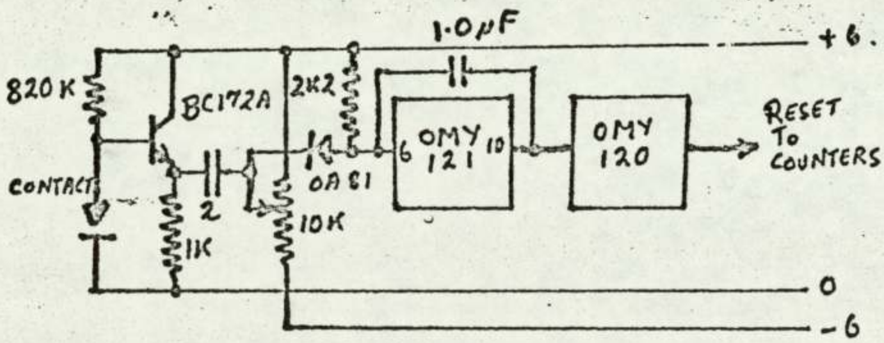


fig. Original End-of-Tape Detector Circuit.

fig. 20.7



Electro-Mechanical End-of-Tape Detector

fig. 20.8

21. THE COMPLETE ASSOCIATING MACHINE

21. THE COMPLETE ASSOCIATING MACHINE

Using the techniques described in the preceding sections, a complete machine of the type described in section 15 was constructed. The overall block diagram of the machine was given in fig. 15.1, and the individual sub-units of the machine were as follows :-

1.	Counters	see section 16.6
2.	Half-Period Pulse Delay	see section 16.7
3.	Two-or-More Detector	see section 17.9
4.	Output Monostable Circuits	see section 18.2
5.	Memory AND Gate	see section 19.1
6.	Memory Record Circuit	see section 20.2
7.	Memory Replay Circuit	see section 20.3
8.	Master Replay Circuit	see section 20.3
9.	End-of-Tape Detector	see section 20.4

All power supplies used in the complete equipment were commercial stabilised units.

The tape-loop recorder used in the tests is shown in fig. 21.1. This was not originally intended for pulse recording, and consequently the heads had a large value of self-inductance. This proved to be a severe limitation, as mentioned in section 20.2.

The inputs to this machine took the form of twenty illuminated push buttons. When any particular

21. THE COMPLETE ASSOCIATING MACHINE

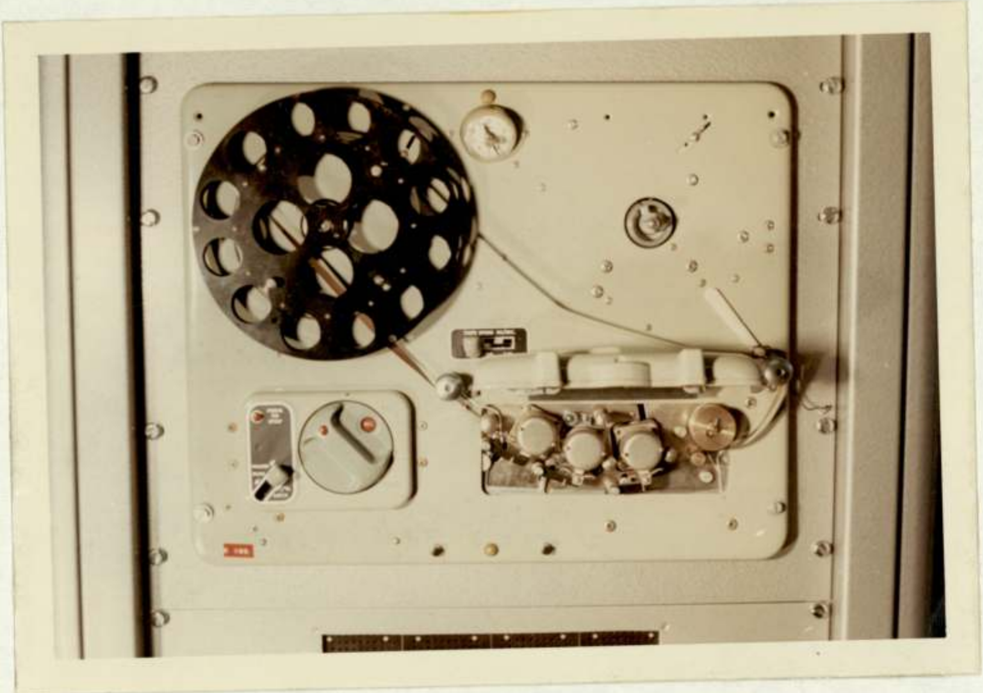


fig. 21.1 THE TAPE-LOOP RECORDER

button was pressed, its internal lamp was illuminated. At the same time, pulses were fed to the two-or-more circuit from the appropriate counter. If two buttons were pressed at the same time, then pulses were passed from the two-or-more detector to the memory record circuit. These pulses were recorded on the memory track of the recording tape loop, at a position corresponding to the correct prime product for the particular pair of push-buttons pressed simultaneously.

Once an association had been recorded at any position on the memory track of the tape, it was possible to make subsequent use of the stored association-information. Pressing one of the buttons previously associated then caused the lamp corresponding to the other button to light.

Typical of the tests carried out with this complete machine is the following :-

Consider four inputs A,B,C,D.

Associations Recorded :-	AB ;	AD ;	BC ;	CD
Input Stimulated :-	A	B	C	D
Ouputs Responding :-	B,D	A,C	B,D	A,C

The major limitation of the machine described is the difficulty of extension caused by the nature of the prime-number system. However, the work described in this thesis led directly to the possiblity of a simplified machine, as discussed in section 22.1.

22. FUTURE WORK

22.1 NEW METHOD OF ASSOCIATION

The method of association using prime numbers which was adopted in the present work evolved for two reasons :-

1. It was derived from the earlier work using prime frequencies.
2. All inputs could be treated in exactly the same way.

However, it has been seen that the method is extremely wasteful in storage facilities. An alternative method is possible which makes a much better use of the storage facilities but in which the inputs must be divided into two mutually exclusive groups. Suppose one group of inputs produce long pulses which persist for ten master pulse periods, while the other group of inputs produce only short pulses each corresponding to one particular unit value. The process is illustrated for the case of 70 and 04 in fig. 22.1 .

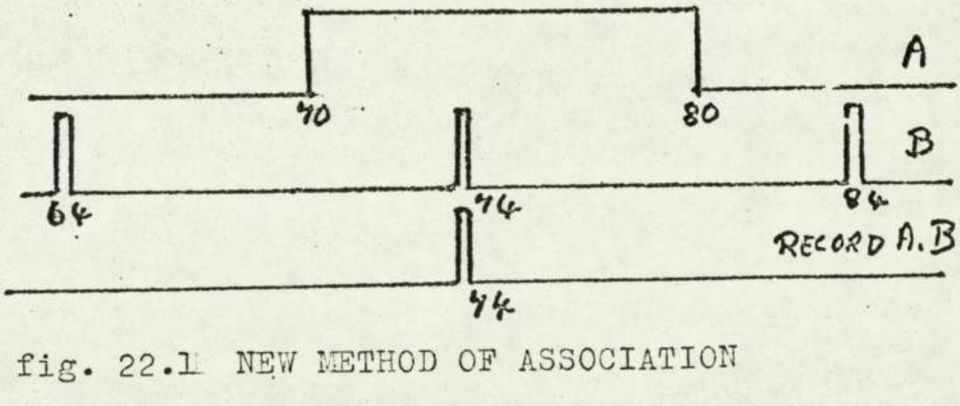


fig. 22.1 NEW METHOD OF ASSOCIATION

At the 74th master pulse there is coincidence and a pulse can be recorded. If then in future the "70" input pulse is present, it coincides with the recorded "74" pulse. The coincidence can be used to produce a "4" output. With such a scheme, the inputs would be divided into two groups, "tens" and "units". There is no reason why the base should necessarily be decimal, and a base such as 12 or 20 or 100 could equally well be used. By increasing the storage facilities by four times, the division into two groups could be eliminated. Each input would then produce one set of "ten" and one set of "unit" pulses. Methods such as this, involving greater complication but improved storage economy, will probably be used in future work.

This new method[38] was discovered some way through the course of the present research, but it was decided to continue with the prime number scheme because of its simpler symmetry, storage space not being a problem in the initial work here described. The new method will have the additional advantage of an improved speed of response. In order to re-introduce a degree of probability, a train of short pulses can be used to replace each long "tens" pulse.

22.2 SEQUENTIAL STIMULI

It is important to note that in the present research, time-sequence storage has not been a fundamental requirement, and that the only consideration has been the simultaneity of stimuli. There are various ways in which time-sequence could be made to affect the operation. One method involves the application of inputs to a delay line having outputs from a number of tappings. These outputs can then be associated with inputs appearing at a later time. In many cases there will in any case be overlap between adjacent stimuli, and between adjacent associations.

It is of interest to consider how time-order affects the operation of a machine in which a certain time-integral of excitation is required to produce a particular output. Suppose that associations X,Y,Z occur in succession and that each produces a certain integral of excitation of a particular output. Suppose that X,Y,Z correspond to inputs AB,BC,CD, ie the inputs A,B,C,D have appeared with overlap in time. Now if the inputs appear with D out of time-sequence as ADBC, then only the association Y=BC appears. The time-integral of excitation of the output in question is consequently much lower than it would be if the inputs appeared in the order ABCD.

It is of interest to consider a complete list of the possible orders of stimulation of four inputs A,B,C,D, and the resulting numbers of overlapping stimuli AB, BC, CD, DC, CB, BA, which occur in each case :-

ABCD	3	DCBA	3
ABDC	2	DCAB	2
ADCB	2	DABC	2
BACD	2	CDBA	2
BADC	2	CDAB	2
BCDA	2	CBAD	2
ACBD	1	DBCA	1
ACDB	1	DBAC	1
ADBC	1	DACB	1
BCAD	1	CBDA	1
BDCA	1	CABD	1
BDAC	0	CADB	0

It is seen from the above table that there is some probability of obtaining sequence sensitivity merely by use of a simple threshold arrangement at the output, combined with time-integration of stimuli. However, some ambiguity is inherent since a sequence and its inverse (e.g. ABCD and DCBA) produce the same result.

22.3 POSSIBILITY OF INTERLACING

In the association-recording method based on the properties of prime numbers, the whole of the possible storage capacity is not used. For example, if the prime number 2 is discarded, then no storage space corresponding to an even number can be occupied by a prime product. The possibility then arises of making use of the even-number spaces to double the storage capacity by adopting an interlaced form of store.

Suppose a store holds an odd number D of digits. Then for any odd prime A, the number (A-D) is not prime for $A - D = \pm 2X$, where X is some number. Similarly, for any prime product AB, the number (AB-D) is not prime since $AB - D = \pm 2X$.
It follows that if a store holds an odd number D of digits, the larger prime products AB might be stored if required by interlacing with smaller odd prime products, so effectively using the memory delay line twice. However, if X itself is a prime product MN, then 2X corresponds to one of the storage positions reserved for product MN. It is best to avoid such problems in the present work, though the possibility of interlacing of recorded information might be considered for later use.

22.4 POSSIBLE USE OF ONLY TWO COUNTERS

It can be assumed that a long train of master pulses is being provided repetitively by a central pulse generator or by a master track on the memory recorder. If only a maximum of two inputs is to be associated at any one time, then in theory only two counters are required. The presence of a particular input, corresponding say to prime number 13 would then set up one counter to count 13. The presence of another input, corresponding say to prime number 17, would set up another counter to count 17.

There would be various difficulties with such a scheme. For example, one must consider what would happen if ever three inputs were energised simultaneously. Another problem is that if the energisation of an input commenced in the middle of the master pulse train, a spurious result would be obtained. Consequently, arrangements would have to be provided to ensure that a counter could only start its counting at the beginning of the master pulse train, and never in the middle of this train. This is a basic requirement.

There is a more serious problem. Suppose that "13" and "17" have been associated in the past and that the precaution above has been observed. Then pulses are recorded on the memory track at positions 221, 442, etc. Now suppose that at some time in the future, input 13 is energised and that it is supplied to an AND gate together with the pulses obtained from the memory track. The AND gate then gives out a pulse. However, there is no indication that this output corresponds to the input 17, and some steps would have to be taken to divide the number 221 by the number 13 to obtain the correct associated 17. One method of achieving this which was considered was the addition of a further counter to count the number of "13" pulses obtained before coincidence was obtained with the recorded "221" pulse.

22.5 LARGER NUMBER OF INPUTS

Suppose that a tape loop having a total capacity of N possible separate recording positions is used. With the prime-number scheme it was shown in section 14.5 that the maximum number of different available channels is given by :-

$$\frac{A\left\{\frac{2}{3}\sqrt{A} - 1\right\}}{\ln A}$$

, where A is the smallest prime number used.

The tape capacity N is required to be approximately equal to A^3, so that the maximum number of different available channels is given by :-

$$\frac{3.\ \sqrt[3]{N}\left\{\frac{2}{3}\ \sqrt[6]{N} - 1\right\}}{\ln N}$$

This quantity is plotted against the value of N in fig. 22.2, the value of N being plotted on a logarithmic scale. It is seen that even for 100 input channels, the required tape storage capacity is becoming rather large.

In order to use larger numbers of inputs with the prime-number scheme, it will be necessary to have a large number of magnetic recording tracks, for example on a magnetic drum. In addition, it might be possible to use interlaced storage as discussed in section 22.3. However, it will perhaps be better to accept the limitations of the method of section 22.1.

This has the advantage of a greatly increased possible storage capacity and a consequent increase of the permissible number of inputs. It will have the additional engineering advantage that it will greatly reduce the amount of counting equipment required.

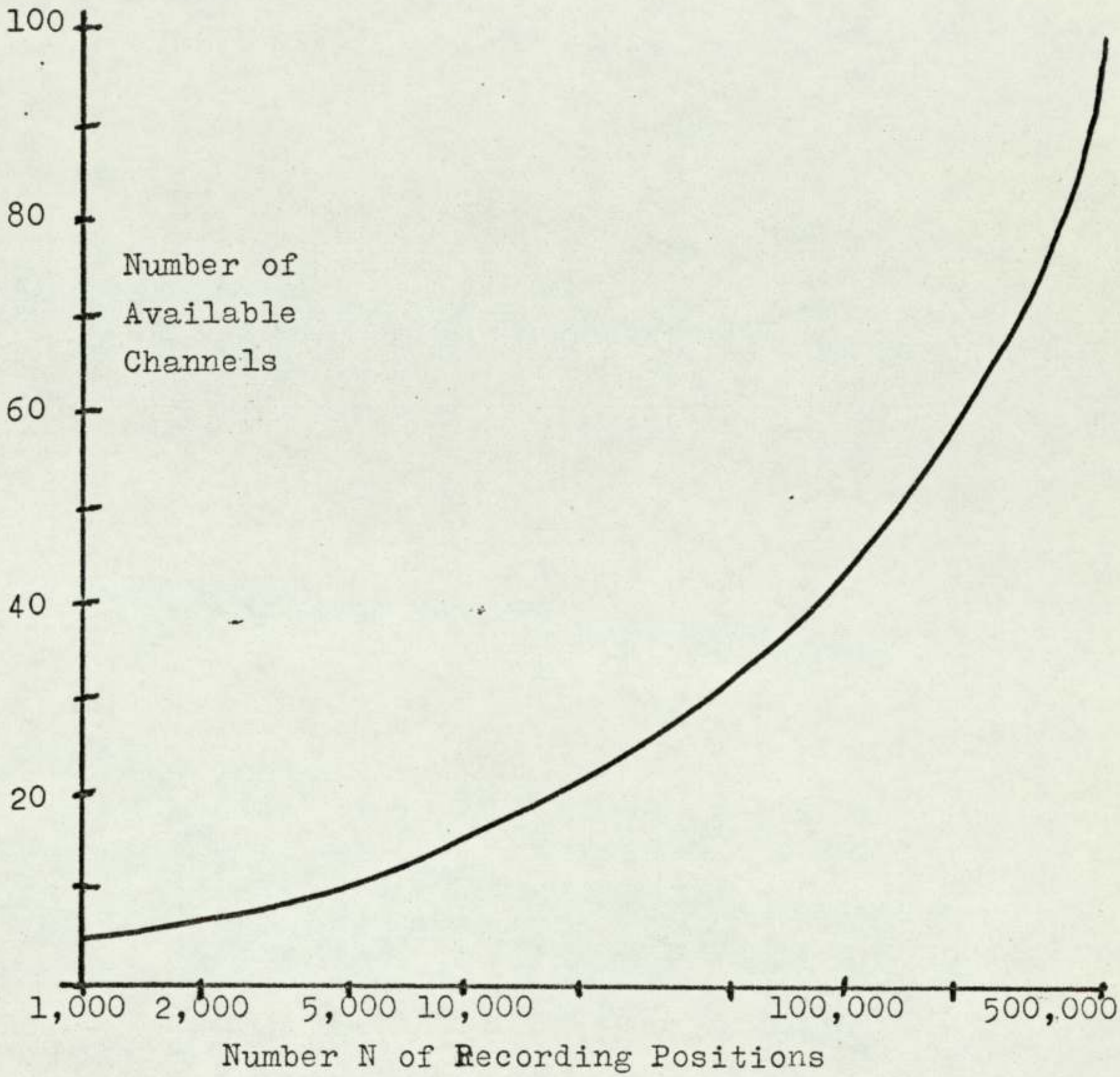


fig. 22.2 Number of Available Channels against Number of Recording Positions on Tape.

22.6 PROBABILISTIC RECORDING

The importance of ensuring the possibility of the later introduction of the feature of probability has been mentioned in section 4. This has been kept in mind throughout the present work, to ensure that any methods adopted did not make impossible the subsequent introduction of probabilistic recording.

It is required that the occurrence of any single association between inputs shall not necessarily saturate the section of memory store reserved for that particular association. Instead, it is required that the more frequently that a particular association between any two inputs has occurred in the past, then the more likely it is in the future that one alone of the two inputs is able to stimulate the output corresponding to the other input.

One way of producing this feature is by the introduction of a random perturbation of the position at which the pulses corresponding to any particular association are recorded on the magnetic memory medium. The more often that a particular association has occurred in the past, then the greater will be the density of pulses recorded around the corresponding tape position and the greater will be the probability that a future pulse occurring in that nominal position will correspond to one of the previously recorded pulses.

The required random perturbation could be introduced, for example, by production of a random variation of the speed of movement of the recording medium, as discussed in section 15.3. Alternatively, electronic methods of random position perturbation could be introduced. Since the latter is likely to be the more controllable method, it is preferred for use in future research.

22.7 FORGETTING

Forgetting is a further feature which, like the probabilistic recording discussed in section 22.6, has been kept in mind constantly during the present work. This was in order to ensure that the possibility of its later introduction would not be ruled out by the nature of the methods adopted.

There are several possible methods by which the forgetting feature might be introduced. One depends on the slow partial decay of information stored on a magnetic tape which can occur when a small high-frequency magnetic erase field is applied to the tape. Repetition of such partial decay leads to an effective gradual decay of stored information. This effect was first noted in the course of the present work, and it is hoped later to carry out further investigations and to determine if it can be used in introducing a forgetting action.

In the final version of the associating-machine, as described in section 21, the information is stored in the form of short pulses on the magnetic medium. Consequently, one possibility which might be used to introduce the feature of forgetting is the imposition of short random erase pulses, applied to one of the magnetic recording heads in an opposite direction to the record pulses. This would then have the effect of introducing random erasure and so of causing a random forgetting of stored information. If the recorded pulses have a random position distribution about the correct points on the recording medium, as discussed in section 22.6, then the erasing pulses should also have a random distribution of position. The forgetting erasure pulses should then take longer to erase completely a memory of any association which has occurred very frequently in the past than to erase the memory of a single past association.

22.8 INHIBITION

A further feature which has been kept in mind for possible introduction at a later stage in the continuing research is the feature of inhibition.[1]

There are two forms of inhibition which must be considered. For one form, it is required that the occurrence of certain inputs should be capable of being associated with the stimulation of definite inhibiting outputs. As an example, in the human arm the tricep muscles are opposed by the bicep muscles, and it is necessary to inhibit the action of one set when the other set are stimulated. This is a quite straight-forward action which can be accomplished by the complete machine described in section 21 without further modification. It would merely be necessary for certain outputs to operate in an inhibitory sense to certain other outputs. This would be accomplished externally to the associating machine.

However, a second form of inhibition may be active at a lower level of nervous activity. Observation of the action of living nerve cells indicates that certain synaptic inputs to some cells appear to be capable of inhibiting the output from that cell.

If this is so, then such inhibition might well be capable of direct reduction of stored memory levels. This is conjectural, though it does appear to agree with observation at a molar level.

It is consequently desirable that the possibility of later introduction of active direct inhibitory association should be catered for, In the present equipment, this would not be difficult. Reverse pulses, such as those mentioned in section 22.7, but only appearing at a definite position on the recording medium, could cause probabilistic erasure of memories previously recorded at that position. It would then only be necessary that, while the occurrence of certain pairs of inputs caused positive polarity associatory pulses to be recorded at a point, the occurrence of some other pairs of inputs could cause negative polarity erasure or inhibitory pulses to be applied at the same point. Thus, this feature can be introduced simply by a reversal of the polarity of some of the "record" pulses.

23. RELATIONSHIP TO OTHER WORK

23. RELATIONSHIP TO OTHER WORK

The action of the TOR circuits of sec. 17 is at first sight similar to that of the threshold logic circuits which have been widely discussed in published work on pattern recognition. Examples are the work of Kamentsky,[48] Rosenblatt,[4] Widrow[49] and of Highleyman.[18] There are many more. To avoid possible confusion, it is therefore as well at this point to note two features of such published work which distinguish it very clearly from the present work.

1. It is an essential feature of these published works that the threshold levels of the various inputs are not all identical.

2. In addition, this weighting of the various inputs is in general carried out manually rather than automatically.

Such proposals have often been backed up by extensive theoretical work, though the practical results obtained are generally less impressive. In much of the published work on adaptive pattern recognition, the aim appears to have been the achievement of eventual permanent error-free operation. In the present work it is recognised that this is not a feature of biological systems.

If any information is to be conveyed by a system of communication, it follows that at the receiver there must be only a limited "a priori" knowledge of and absolutely no control over the transmitted input signal.[57] The judgment required at the receiver involves statistical decisions.

Some of the published theoretical work on character recognition has been based on the use of statistical decision functions.[14] It is important to note that much of this work has been concerned only with the process of recognition, rather than with the process of learning to recognise. A critical part is played in the formulation and application of decision theory by "a priori" probabilities. When such probabilities do not exist, then such a theory is inapplicable.

It has been recognised that the "a priori" probabilities are unknown to the designer of a system. However instead of simply accepting the inapplicability of decision theory in such a case, it has been usual to place restrictions on the systems considered in order to make them fit the theory. In addition, the optimum theoretical decision functions have been too difficult to implement and degraded forms have been used.[19] The "linear decision function" elements have been in effect majority logic elements as described earlier.

Some of the theoretical work which has been published uses a multi-dimensional representation of the input conditions[12]. With this approach it is assumed that it is necessary to consider every possible combination of all inputs, present or absent. The receptor output for a particular input pattern is a set of numbers defining the coordinates of a vector-tip in an N-dimensional measurement space. Each vector is in effect an association between all inputs, whether present or absent. For example :-

Inputs :-	A	B	C	D	E	F	G	H	Outputs
	0	0	1	0	1	0	0	1	X
	0	0	1	1	0	0	0	1	Y

The adoption of this form of representation in a practical device would necessitate the use of a great deal of equipment if more than a few inputs are to be dealt with. There are 2^N different states of N inputs. Not only would an astronomical amount of storage space be required for practical use of such a full set of vectors, but much of the space would be wasted since many of the particular combinations of inputs would be extremely unlikely to occur.

Caution is required when considering much of the published work on character recognition. Partly this is because most of it has considered not machines which learn, but only machines which can be set up manually to differentiate between characters.[50] It is stated by many writers on the subject[51] that pattern recognition machines have to be "taught". It is important to take great care with the use of language in this subject, and it is therefore desirable to reword this sentence by using the words "set-up" to replace "taught". This removes any implications of co-operative activity on the part of the machine. Now the present work has been directed at the design of machines which do not have to be set up, but which can inherently learn to associate certain events with certain other events which occur in the environment of the machine.

Another point is worthy of mention. Devices based on the application of decision theory will have a response which is exactly correct most of the time, but whenever an error is made at all it will be a large error. Such a response would probably be fatal in a living organism. There would be no learning about the effects of slow changes in the environment. More important, because of the occasional large errors, natural selection could be expected to operate decisively against this form of response.

Devices based on the application of filter theory,[52] are intended to minimise the mean square difference between the actual response and the desired response.

Such devices are intended to discriminate continuously against large errors, though at the expense of continual small errors. Such a form of response is ideal for a machine which must learn and which must display a measure of homeostasis. In general methods based on the application of decision theory are suitable for use where the states of the desired signal can be exactly defined a priori. However, if the states of the desired signal are ill defined (e.g. speech waveforms), then filter methods are better.

In work on visual pattern recognition or character recognition, the design of the original characters is always based on the necessity for them to be recognised by a human as well as by a machine. (The writer has pointed out[5] that in some cases, for example cheques, this feature is quite unnecessary; the human can read the information elsewhere on the same document.)

Particularly in work on speech, it is necessary to recognise that the signals acceptable to humans are extremely redundant. The minimum standards acceptable as recognisable by a machine might be much reduced if only it was known what are the important components of a speech signal.

The research described in the present thesis has led to a simulation of some aspects of the operation of the nervous system which meets the requirements laid down by MacKay[56]. These are :-

"..a statistical model in which connexions between elements (1) are incompletely specified (2) function with adjustable probability, and (3) grow in complexity step by step with the development of internal organisation to match the structure of the environment."

24. CONCLUSIONS

24. CONCLUSIONS

Separate conclusions have been presented earlier on the methods of :-

Sinusoidal Re-Recording (section 7.7)

Pulse-width Modulation (section 8.8)

and The Short-Pulse Method (section 14.7)

Consequently, these will not be repeated here.

One aim of the present work has been to devise general and adaptable cybernetic engineering methods which can be used in the investigation of the effects of Pavlovian conditioning in machines.

The methods arrived at are adaptable for trial with many different basic possibilities. Examples are :- multi-way associations ; hierarchies of associations ; inhibitory effects. As required, the methods are extendable for use with many inputs.

With the method of non-dynamic circulation which was finally adopted, only one of the basic requirements for an associatory machine, as presented in sections 3 and 4, has not yet been achieved. This is the requirement for probabilistic operation.

However, possible methods for the later introduction of this feature, together with the associated features of inhibition and forgetting which were mentioned in section 4, have been suggested in section 22. The requirements of small size of the basic associatory-information store and of exclusive recording of associations between inputs have been met by the complete machine described in section 21. The research work leading to the production of this machine has not only verified the practicability of the method, but it has led directly to the new method described in section 22.1. [38]

It is probable that future work in this field will make use of this new method, since the new scheme can give an improved economy while retaining most of the features of the arrangement described in section 21.

With this new method, it will be possible economically to build extendable machines capable of handling many hundreds of thousands of inputs and of outputs, without requiring excessive amounts of space or of equipment. With these devices, investigations of many various possible methods of organisation of learning machines will be practicable.

Thus, the research described in this thesis has produced valuable knowledge and techniques for use in future cybernetic work.

25. REFERENCES

25. REFERENCES

1. Young, J.F., Cybernetics, Iliffe, 1969.

2. Sklansky, J., Adaptation, Learning, Self-Repair and Feedback, I.E.E.E. Spectrum, V 1, p 172, 1964.

3. Nagy, G., State of the Art in Pattern Recognition, Proc.I.E.E.E., V 56, p 836, 1968.

4. Rosenblatt, F., Principles of Neurodynamics, Spartan, 1961.

5. Young, J.F., Cybernetics, British Communications and Electronics, V 12, p 302, 1965,

6. Descartes, R., A Discourse on Method, Dent, 1965.

7. Pavlov, I.P., Conditioned Reflexes, Dover, 1960.

8. Walter, W.G., The Living Brain, Duckworth, 1953.

9. Zemanek, H., Kretz, H., and Angyan, A.J., A Model for Neurophysiological Functions, Information Theory, 1960, (ed. C. Cherry,) p 270, Butterworths, 1961.

10. Huggins, P., Two Experimental Learning Machines, J.I.E.E., V 6, p 702, 1960.

11. Angyan, A.J., Machina Reproducatrix, Mech. of Thought Processes, P 933, H.M.S.O., 1959.

12. Steinbuch, K., and Piske, U.A.W., Learning Machines and their Applications, Trans.I.E.E.E., V EC 12, p 846, 1963.

13. Underwood, B.J., Forgetting, Sc.Amer., V 210, p 91, 1964.

14. Chow, C.K., An Optimum Character Recognition System Using Decision Functions, Trans.I.R.E., V EC 6, p 247, 1957.

15. McCulloch, W.S., and Pitts, W.A., A Logical Calculus of the Ideas Immanent in Nervous Activity, Bull.Math.Biophys., V 5, p 115, 1943.

16. Beurle, R.L., Properties of a Mass of Cells Capable of Regenerating Pulses, Phil.Trans.Roy.Soc., V 240B, p 55, 1956.

17. Uttley, A.M., The Transmission of Information and the Effect of Local Feedback in Theoretical and Neural Systems, Brain Research, V 2, p 21, 1966.

18. Highleyman, W.H., Linear Decision Functions, with Application to Pattern Recognition, Proc.I.R.E., V 50, p 1501, 1962.

19. Griffin, J.S., King, J.H., and Tunis, C.J., Pattern Identification Using Adaptive Linear Decision Functions, I.E.E.E. Wescon., Paper 6.4, 1963.

20. Young, J.F., Electronic Relays for Power System Protection : Possible Application of the Hall Effect, Electrical Rev., V 179, p 352, 1966.

21. Carmichael, R.D., and Smith, E.R., Mathematical Tables and Formulae, Dover, 1962.

32. Crane, H.D., The Neuristor,
Proc.I.R.E., V 50, p 2048, 1962.

33. Yoshizawa, S., et al, An Active Pulse Transmission Line,
Electr. and Commun. in Japan, V 46, p 22, 1963.

34. Cote, A.J., A Neuristor Prototype,
Proc.I.R.E., V 49, p 1430, 1961.

35. Rosengreen, A., Experimental Neuristor,
Electronics, V 36, p 25, 1963.

36. Steinbuch, K., and Piske, U.A.W.,
Learning Matrices and their Applications,
Trans.I.E.E.E., V EC 12, p 846, 1963.

37. Verzeano, M., and Negishi, K., Neuronal Activity in Cortical and Thalamic Networks,
J.Gen.Physiol., V 43, p 177, 1960.

38. Young, J.F., and N.R.D.C., British Provisional Patent no. 30657/68.

39. Ball, W.W.R., History of Mathematics, Macmillan, 1924.

40. Young, J.F., A Transistor Characteristic Curve Tracer, Electronic Engg., V 31, p 330, 1959.

41. Young, J.F., Coupling Dekatrons with High-Voltage Transistors, Electronic Engg., V 39, p 323, 1967.

42. Young, J.F., Transistor-Diode Static Switching Units, Electronic Engg., V 34, p 595, 1962.

43. Young, J.F., Counters and Shift Registers Using Static Switching Units, Electronic Engg., V 37, p 161, 1965.

44. Young, J.F., Cyclic Binary-Decimal Complementing Codes, Control, V 8, p 226, 1964.

45. Young, J.F., G.E.C. Internal Report SWD

46. Young, J.F., Applied Electronics, Iliffe, 1968.

47. Hoagland, A.A., Digital Magnetic Recording, Wiley, 1963.

48. Kamentsky, L.A., and Liu, C.N., Computer Automat Design of Recognition Logic, I.B.M.J., V 7, p 2, 19

49. Widrow, B., et al., Practical Applications for Adaptive Data-Processing Systems, I.E.E.E. Wescon Pt. 7, paper 11.4, 1963.

50. Sklansky, J., Learning Systems for Automatic Control, Trans.I.E.E.E., V AC11, p 6, 1966.

51. Rosenblatt, F., Perceptron Simulation Experiments Proc.I.R.E., V 48, p 301, 1960.

52. Hawkins, J.K., Self-Organizing Systems, Proc.I.R.E., V 49, p 31, 1961.

53. Williams, F.C., and West, J.C., The Position Synchronisation of a Rotating Drum, Proc.I.E.E., V 98, pt.2, p 29, 1951.

54. Booth, A.D., A Magnetic Digital Storage System, Electronic Engng., V 21, p 234, 1949, July.

55. Taylor, W.K., Pattern Recognition by means of Automatic Analogue Apparatus, Proc.I.E.E., V 106B, p 198, 1959.

56. MacKay, D.M., Modelling of Large-Scale Nervous Activity, S.E.B. Symposium No. 14, p 192, 1960.

57. Young, J.F., Information Theory, to be published by Butterworths.

26. ACKNOWLEDGEMENTS

26. ACKNOWLEDGMENTS

I wish to record my thanks to my Research Supervisor and Head of Department, Professor J.E. Flood . I am particularly indebted to him for ensuring that adequate laboratory facilities and measuring equipment were available for this work.

I also wish to record my thanks to the Science Research Council for providing a Research Grant without which this work would not have been possible.

I also wish to record my thanks to Mr. Norman Hollingshead, the S.R.C. Research Technician employed on this project, for his unswerving loyalty. He has been responsible for all of the constructional work on this project. In addition, the research has been furthered by his intense interest and by his scientific attitude.

I also wish to record my thanks to the National Research Development Corporation for dealing with the patent protection of this work.

I also wish to record my thanks to my students, for stimulating discussions on many aspects of this work.