THE UNIVERSITY *of* EDINBURGH

MODELS OF DISTRIBUTED ASSOCIATIVE MEMORY

DAVID WILLSHAW

Ph.D. University of Edinburgh 1971

CONTENTS

ABSTRACT

This work is concerned with the problem of constructing associative memory models which store and retrieve information in a distributed fashion - that is each store location can be involved in the storage of more than one piece of information. The models are to be logically simple in form and efficient in their use of store. It is hoped that those proposed may find a place in neurophysiological theories of memory.

From the starting point of seeking a less complex representation of a holographic model of memory called the Holophone, several memory models have been proposed and their properties investigated by analytical and computer simulation methods. All the systems considered belong to a family of memory models which are described by a pair of matrix equations.

It is found that the models which perform well are non-linear devices which have to store and retrieve highly redundant information. One of these, the Associative Net, is closely related to a recent theory of the cerebellum.

Some of the contents of Chapters 3, 4 and 5 have already appeared in papers published in collaboration with H.C. Longuet-Higgins and O.P. Buneman. Otherwise this dissertation is believed to be original in content and / . .

and in arrangement.

This work was carried out under the supervision of Professor H.C. Longuet-Higgins. The author is very grateful to him for his sympathetic guidance.

NOTES ON NOTATION

We are going to discuss the storage and retrieval of pieces of information. We call these "signals", "patterns" or "messages". "Signals" are stored in temporal models such as the Holophone, "patterns" in optical models (the Correlograph) and "messages" are stored in network models like the Associative Net. We have also endeavoured to employ the term "message" in more general contexts, but it should be recognised that these three terms are in effect synonymous.

We represent the information about one of a number of messages by the values of column vectors of a matrix. $X^{(n)}$ is the nth column vector of matrix X. Only in Chapters 2 and 8 is it necessary to distinguish vectors from arithmetical quantities. We do this by underlining. For example, the vector $X^{(n)}$ becomes $\underline{X}^{(n)}$.

The Associative Net processes messages, denoted by vectors with components of value 1 or 0, which we imagine are represented by a distribution of pulses sent along the Net's lines. Thus a particular active line transmits a pulse or a "1". Similarly, an inactive line is identified with a "0".

CHAPTER 1

Introduction

1.1 The Problem

We consider the problem of constructing simple devices to store and retrieve information reliably. We constrain our problem by requiring that the systems be distributed (or non-local) associative stores, functioning in parallel, which have to deal efficiently with a small number of messages drawn from a very much larger ensemble. The emphasis is placed on distributed memories of this type because to the author's knowledge they have not been the subject of detailed analysis before and because it is hoped that they may have some relevance to biological theories of memory. The models are required to be efficient in their use of store, that is their performance in the storage and retrieval of information should not be much worse than the best that can be expected on information theoretical grounds.

1.2 Definitions

A local store is one in which each store location contributes to the storage of one and only one message. In a distributed system each location may be identified with more than one stored message.

We shall define the word associative in terms of the concept of Classical Conditioning.

Consider / . .

Consider a system in which the event X is causally related to the event Y which follows it. If the event Z occurs coincidentally with X and a mechanism operates, such that subsequently Z on its own will cause Y, then Classical Conditioning is said to have taken place. We employ this analogy by considering the special case when X and Y are identical. Then information has been stored by the mechanism of Classical Conditioning when the information to be stored (Y) and the address to locate it (Z) are together input to the store, so that subsequently the address alone will locate the stored information. This is an <u>associative</u> store.

This mechanism is to be contrasted with that of Operant Conditioning. Here, consider events Z and Y. If on the occasions when Z occurs followed by Y, occurrence of the event Y causes the Z-Y link to be strengthened, than the strong causal relationship which is ultimately set up between Z and Y is said to have been caused by Operant Conditioning.

1.3 Local memories

We will first look at an example of a local store, namely a conventional computer memory, for which many sophisticated storage schemes have been developed.

Consider a number of messages and their associated addresses. Each address and each message is represented by / . .

by a string of binary digits and we suppose that the address bit-strings are all of length N. We will give the system the task of storing an arbitrary selection of the 2^N different possible messages, each identified by a unique address bit-string. If it is known that about 2^N (but no more than 2^N) messages are to be stored then it is a simple matter to identify each possible address bit-string with a different location, where the message string corresponding to that address can be stored, and from whence it is subsequently read out. In this situation the number of bits stored per binary register is not very much less than one bit per register, which is the information theoretical maximum.

However, if, as is usually the case, the number of messages to be stored is much less than the 2^N possible, then storage space is wasted by using this method. What is usually done, therefore, is to assign by a well-defined procedure one of a small number of new addresses to each of the 2^N possible addresses. This small number is usually taken to be not less than the total number of messages to be stored, an upper limit to which must be known in advance. Each reduced address is now identified with a particular location of an area of store, now known as the hash table, and each message bit-string is stored in the hash location corresponding to its reduced address. The difficulty now is that

more / . .

more than one message will be given the same reduced address. In this case, the messages are put into an overflow store, whose address is placed in the appropriate entry of the hash table. Retrieval of a message from an overflow store usually entails a comparison, address by address, to find the message whose N-bit address is identical to the N-bit address presented. This method has several variations. There are different ways of identifying each original address with a reduced address and different ways of organising the overflow store (Hopgood, 1969). As far as the latter is concerned, if speed of operation is required then, as the need arises, each hash location is given its own overflow store, which occupies an area distinct from the hash table. Conversely, economy of storage space is produced at the expense of speed of execution by employing unused hash locations as the overflow store. This variation enables the theoretical maximum efficiency of storing information to be attained.

1.4 Distributed memories

The hypothesis that living organisms may store and retrieve information non-locally has existed for many years, and it has been even more in prominence recently, owing to the invention of a potential distributed store called the hologram, which has led some people to draw elaborate parallels between biological memories and holographic / . .

holographic plates. We will consider the more fundamental problem of constructing simple distributed memories which will deal efficiently with the non-trivial problems that arise in the storage and retrieval of information.

We have seen that a conventional computer store, which is our example of a local memory, can be organised so that it may perform the task we have described efficiently. It is not obvious, however, how a distributed store can be made to perform this relatively simple exercise, especially as it is in the very nature of such a store that locations are shared, so that the record of one piece of stored information may distort that of another. Furthermore, it is desirable that a neurophysiological memory - local or distributed - be able to cope with such problems as retrieving a message from store when using a slightly inaccurate address, or performing efficiently in the face of damage to some of its store locations. A computer store could overcome this first problem, which can be related to that of designing systems able to perform perceptual generalisation, by providing pointers to the correct locations at the inaccurate locations (if they are known), and the second by storing the same piece of information at several different locations, but at the expense of severely decreasing / . .

decreasing the number of locations available for storage. These solutions are not claimed to be the best available, but it is clear that these problems are not ones with which computer stores, and in fact local stores in general, are able to deal with satisfactorily. There is the even more difficult question of designing content addressable stores, to which, incidentally, great effort has been devoted in the field of computer hardware design. Such a system is one in which the address to locate a piece of stored information is itself an arbitrarily selected part of that stored information.

1.5 Objectives

We have attempted to design logically simple systems with the following considerations in mind:

i) A number of pairs of messages, which are drawn from a large ensemble of possible pairs, are to be stored non-locally, so that with the aid of one member of the pair (the address or cue), the other is able to be reproduced accurately from store.

ii) The systems should be able to function in the face of damage.

iii) The systems should be efficient in their use of storage space. It is one problem to construct a distributed memory store; it is a more difficult problem to construct such a system whose performance does not fall / . .

fall short of the best that can be expected from information theoretical considerations. The second problem has been tackled.

iv) The systems should be designed with an eye on biological plausibility. The author is not competent to discuss in detail what may constitute a good distributed model of biological memory but it is hoped that the models discussed may have some relevance to theories constructed on the neurophysiological level.

1.6 The linear/non-linear classification of the models

The class of models considered can be represented by a pair of matrix equations relating the output α (retrieved information) to the input β (address) of one such system

$$\alpha = M^{-1}DM\beta \quad \ldots\ldots\ldots 1.1$$

and

$$\alpha = [[M^{-1}DM]\beta] \quad \ldots\ldots\ldots 1.2$$

The brackets [ ] denote that a non-linear operation has been performed on the matrix that the brackets enclose, that is the value of a component $[Z]_{ij}$ of the matrix $[Z]$ is a non-linear function of the value of the component Z_{ij} of the matrix Z. Each model considered is either <u>linear</u>, in the sense that its input-output relationship can be represented by equation 1.1, or <u>non-linear</u> (equation 1.2). Both equations represent a three-stage process, since the output α is produced from the address β by three successive matrix multiplications. Here M is the / . .

the discrete Fourier Transform matrix. The matrix D contains a record of the stored information, and represents a distributed store, since for each model the method of assigning values to the components of D is such that the value of any one component is determined by the form of more than one stored pair of messages. We see that for models described by equation 1.2, one non-linear operation is involved in storage and one in retrieval.

1.7 Outline

Here is a brief summary of the contents of the following chapters.

It is hoped that the models proposed may have some relevance to living organisms on the neurophysiological level. Consequently Chapter 2 presents an account of some biological theories of memory.

This work had as its origin the analysis of the properties of the Holophone, which is a linear (equation 1.1) frequency analysing system designed to be a temporal analogue of the hologram. The relationship between the Holophone and holography, the mathematical analysis and the supporting computer simulation results of the Holophone's performance as a memory are the subject of Chapter 3. It was found that if the Holophone stores segments of random noise, parts of one segment being used as the address to recover the remainder, then its performance, measured by the magnitude of the signal-to-noise / . .

noise ratio of recalled fragments, is poor. A related linear model, the Off-diagonal Holophone, which differs from the Holophone in the form of the memory matrix D, was analysed and found to perform the same task more reliably.

However, since the form of equation 1.1 describing the functioning of the Holophone is relatively simple, it was thought that there might be structurally simpler models, equivalent in function to the Holophone, which would be more amenable to analysis and would, moreover, be more plausible on neurophysiological grounds than this complex frequency analysing device. Chapter 4 describes such a model. In its non-linear form, called the Correlograph, analytical statements can be made about its performance. It was found that if it is allowed to store highly redundant messages then it can be made to work not far short of the optimum set by information theoretical considerations. A structurally even simpler non-linear model, called the Associative Net (which is related to the Off-diagonal Holophone), is also described in Chapter 4. Similar statements to those about efficiency of storage in the Correlograph apply to the Associative Net.

Chapters 5 and 6 describe the properties of these two non-linear models. In particular, it is shown that, at the expense of sacrificing information storage efficiency / . .

efficiency, they can perform satisfactorily when required to retrieve information given a slightly incorrect address or in the face of damage to the store. Some of the properties of such systems when equipped with feedback loops are also considered and demonstrated by means of computer simulation experiments.

The last two models are discussed in Chapter 7. The non-linear Symmetrical Associative Net was explicitly designed to make use of information about the stored messages which the Associative Net ignores. It is shown that it can store irredundant messages quite efficiently and can be adapted easily in the face of damage to the store. It has the disadvantage, however, that once a certain set of messages has been stored, no more can be added. A related model, the Weighted Symmetrical Associative Net, which was designed to overcome these difficulties, is found to perform less efficiently. The similarity between this model and the Off-diagonal Holophone leads us to re-express the earlier signal-to-noise calculations for this and the Holophone in terms of storage efficiency, and the conclusions of Chapter 3 concerning these models' performance are reinforced.

Chapter 8 is concerned with related work. In particular, the concepts of Classical and Operant Conditioning are used to illustrate the difference between / . .

between the models we consider and a simple Perceptron.

The more speculative questions of biological implications are entered into in Chapter 9. It is found that there is a strong relationship between the Associative Net and a recent theory of the cerebellum.

In Chapter 10 we summarise the formal mathematical relationships between our models and conclude by suggesting that it has been demonstrated that distributed models of memory which are simple in structure and efficient in performance can be constructed. The models which we have found to perform well are non-linear in nature, store highly redundant messages and it seems that at least one of them (the Associative Net) may have neurophysiological relevance.

CHAPTER 2

Background

2.1 Introduction

We will first place this work in its biological context. We will make explicit the concepts of memory and learning and consider local and distributed theories of memory.

The memory of an organism is that part of it which records past events and to which reference is made in the organism's future behaviour. Learning is the manner in which the memory is used. No claims are made for the originality of these statements, but they are included for the sake of avoiding confusion.

A scientific theory of memory will be able to predict how an organism uses previously stored information in its future actions. To arrive at a satisfactory theory it might be good enough to simply observe the organism's input and its output. On the other hand it may be necessary to observe the behaviour of every one of the molecules which make up the organism. Finally a middle course may be acceptable. Here it is hoped that the models proposed will have some relevance to neurophysiological theories of memory, where the nerve cell may be treated as the functional unit and the messages with which it deals are represented as electrical pulses transmitted / . .

transmitted along the nerve fibres. Although it is tempting to embark on rationalisation, it must be emphasized that this is just one of many points of view and is not backed by unshakeable evidence.

On the neurophysiological level, a local theory is one which demands that each synapse be identified with the storage of one and only one piece of information, while in a distributed theory each synapse can be identified with the storage of more than one piece of information.

2.2 Local theories

This type of theory has a very long pedigree. Plato (429 - 348 B.C.) drew the analogy between memory and a block of wax "... whenever we wish to remember, let us stamp the perception or thought upon the wax, as if we were making an impression with the seal of a ring. ... their impressions being clear and distinct and speedily disposed without jostle or confusion each in its proper place are called real and those who have them are called wise." (Plato, The Theaetetus). We also mention Associationist philosophers, such as James Mill (1773-1836) who held that the human mind concerns itself with the linking together of pieces of sensory experience. The British psychologist Alexander Bain (1818-1903) has a more concrete physiological explanation of memory. "for every act of memory, every exercise of bodily attitudes, every / . .

every habit, recollection, train of ideas, there is a specific grouping or coordination of sensations and movements by virtue of specific growth in the cell junctions." He also notes that "there is no improbability in supposing an independent nervous track for each separate acquisition". (Gomulicki, 1953).

A simple formulation of a local theory is that the memory is working as a switchboard with a number of input lines, each linked to the appropriate output line. Learning involves the setting up of new connections. This concept is seen in the doctrine of Connexionism due to Thorndike (1874-1949). "All psychological processes consist of the functioning of native and acquired connexions between situations and responses" (Thorndike,1949), while Pavlov (1849-1936) was not the first to suppose that memory traces and reflex arcs functioned according to similar principles (Pavlov, 1927). He thought of the cortex as comprising many stimulation foci (centres of afferent stimulation), with well established pathways connecting those foci which had become associated by conditioning. More recently, Young (1966, 1970) has suggested that the memory of the octopus consists of a number of simple components, each of which records the consequences of stimulating a particular classifying cell which responds to a particular type of visual or tactile / . .

tactile input. Initially, each classifying cell can respond in more than one way to stimulation. Learning takes place by the inhibition of responses potentially harmful to the animal.

2.4 Distributed theories

What local theories have difficulty in explaining are the facts of perceptual generalisation and the suggestion that stored memories are not destroyed by local damage.

Gestalt Theory

Psychologists of the Gestalt school explained the former in terms of patterns of excitation (Koffka 1935, Köhler 1940). Consider how they would explain the perception of a circle. The sensory input is transformed into a pattern of excitation in the brain, modifying its ongoing activity. That the circle has been seen is noted by the pattern of excitation leaving a record of some description in the brain. When the circle comes into the organism's field of view again, the brain recognises that the current pattern of excitation caused by the circle is similar to the pattern which laid down the original trace and the organism remembers that it has seen the circle before. Perceptual generalisation is now accomplished as follows. A rat will look behind a small door to find food, having previously learnt to look for / . .

for food behind a large door of similar shape, because the patterns of excitation in the brain due to the small door and due to the large door have similar properties which its brain can recognise. This is a distributed theory, as each piece of information is not stored in a unique location. It can be seen, however, that the details of the mechanism underlying the Gestalt theory are not clear. By being incomplete such theories as this must be regarded as being unsatisfactory.

Localised damage

Lashley (1890-1958) although anticipated by, amongst others, Flourens (1794-1867) and Loeb (1859-1924), is the most well known proponent of the view that memory traces are not localized in the cerebral cortex. He found that rats who, after learning mazes, subsequently underwent brain surgery, suffered brain defects dependent on the amount and not the location of brain tissue removed. He inferred that all parts of the cerebral cortex play an equal role in the memory process. One of his theories (Lashley, 1942) (which he subsequently rejected) was that each learned event is represented by a particular pattern of vibrations in the brain. Like all theories of memory depending on the maintenance of electrical activity in the brain (see, for example, Rashevsky, 1938), this does not explain, for example, how human memories survive grand mal epileptic attacks, or how brain activity can / . .

can be severely reduced by cooling or by anaesthesia without serious impairment of memory.

Of the theories which set out to explain Lashley's findings, we will mention that of Roy (1960,1962). He suggested a distributed nerve net with one input and one output channel, made up of a number of identical nerve-cell like units functioning as delay lines. By the mechanism of threshold lowering within the units, the net was able to recall a particular part of a stored signal by using the preceding part as the address. Although terms such as 'threshold' and 'synapse' are employed, the properties of his artificial units do not correspond to what is known about nerve cells. In fact Roy pointed out that his units were not intended to model strictly real neurons or aggregates of neurons.

Ensembles of nerve cells

The concept that a biological memory may not merely comprise a set of disconnected lumps is reflected in the work on the analysis of the properties of ensembles of interacting nerve cells.

In his theory of cell assemblies Hebb (1949), in an attempt to reconcile "switchboard" and "field" theories (e.g. Gestalt theory), put forward the idea of modifiable excitatory synapses - that is the excitatory synapse between an axon and a dendrite is facilitated if activity in the axon coincides with depolarisation of the dendrite.

As / . .

As learning proceeds, cells are modified by this means and they form themselves into interacting groups, called assemblies, each capable of supporting patterns of excitation. One nerve cell can belong to more than one assembly and can change allegiance from one assembly to another. Thus one cell can contribute to the storage of more than one message. Milner (1957) extended Hebb's treatment to include inhibitory synapses and both theories were tested by computer simulation by Rochester and associates (Rochester et al.1956). They found that cell assemblies could only be produced in a set of interacting neuron-like elements if both inhibitory and excitatory synapses were present.

Cragg and Temperley (1954) suggested that an interacting ensemble of neurons be represented as a collection of magnets. They presented evidence of neurophysiological analogies for the properties of the model and, with reference to memory, they suggested how Lashley's results could be explained.

There have been many other attempts made to investigate the properties of an ensemble of neurons, in particular those of stability. Beurle (1956) studied the mathematics of wave propagation through a set of neuron-like elements, which he treated as a continuous system, whose properties were based on the cytological evidence of Sholl. By using the hypothesis that cells which

which fire have their threshold lowered so that they become easier to excite in the future, he did suggest how associative learning could take place.

Influence of Holography

The recent interest in distributed memories can be traced to the technical advances in holography during the last few years.

Holography is a method of information storage employing coherent beams of electromagnetic radiation. This was invented by Gabor (1948,1949,1951) and has achieved technical importance with the arrival of the laser (Leith and Upatnieks 1962, Stroke 1966). The holographic principle will be described by reference to an experiment. A coherent beam of light from a laser is split by a half-silvered mirror. One beam A passes through a photographic transparency T of a circle, and strikes a photographic plate, where an interference pattern is formed between it and the other beam B which is incident directly onto the plate. Each point on the plate receives light from each point on the transparency. The plate is developed and a positive transparency of it, the hologram is put in its place. When this is illuminated by the beam B in the absence of the original transparency, an image of the circle is still seen by looking through the hologram in the direction of the original position of T. In principle, by exposing the same photographic plate / . .

plate in this way to more than one pair of beams of light, the hologram can be made to act as a distributed store.

In the example given above, which is an illustration of Fourier holography, concerned with the recording of two dimensional images, the beam of light is acting as the address to locate the particular piece of information required.

Van Heerden (1963b) seems to have been the first person to draw the analogy between holograms and distributed theories of memory. He discussed the similarities between optical information storage in solids by using coherent light and Beurle's suggestions as to how associative learning could take place in a nerve net by means of modifiable thresholds. Van Heerden pointed out that such systems are able to store large amounts of information and he stressed the necessity for a calibrating system of pulses in the brain in order to maintain exact phase relations between waves. Other people have discussed the importance of holography in its relationship to Lashley's experimental findings, and some have produced holographic brain models (Pribram 1966,1969; Westlake 1968).

2.5 Conclusion

These last remarks about holography conclude our brief / . .

brief survey of theories of memory. The starting point for this work did, in fact, lie in holography for, as we shall see in Chapter 3, we commenced our studies of models of distributed memory by analysing the properties of a holographic type memory called the Holophone.

CHAPTER 3

Holographic Models

The Holophone and the Off-diagonal Holophone

3.1 Introduction

This chapter will aim to describe the functioning and the mathematics relevant to its performance in storing and retrieving information of a holographic-type memory called the Holophone.

The Holophone came into being by way of holography; for that reason a summary of some aspects of holography will be made. As we have mentioned in Chapter 1 it can also be viewed as one of a number of distributed models which are related to each other by the similar mathematical equations which describe their behaviour. Another member of this family of models will be discussed when the principles underlying the Holophone have been put forward.

3.2 Holography

Consider a number of electromagnetic wavefronts, bearing specific phase relationships one to another, which interfere with each other in a particular region of space, with the result of modifying the transmittance of a detector placed there. The record of the interference pattern so made is called the hologram. Subsequently, when some of the wave fronts which took part in the recording process are sent through the hologram the/ . .

the remaining wave fronts are reconstructed.

What is happening here can be formulated mathematically by considering the following example (Collier 1966, Stroke 1966).

Two objects A and B are illuminated by monochromatic coherent light from a laser by means of a split beam arrangement (Although these remarks refer to optical systems, holograms can in principle be constructed using electromagnetic waves of any frequency.). Light is reflected diffusely at the surface of the objects and interferes in a photo-sensitive solid whose transmittance at any point $\underline{r}$ in it is assumed to be linear in intensity - that is it is assumed to change in proportion to the intensity of light at that point. (Figure 1A on page25). Omitting time variation factors, if the complex amplitudes of the waves diffracted from objects A and B at any point $\underline{r}$ in the solid are $F_A(\underline{r})$ and $F_B(\underline{r})$ then the change in transmittance at that point is

$$\Delta t(\underline{r}) = \lambda(F_A(\underline{r}) + F_B(\underline{r}))(F_B^*(\underline{r}) + F_B^*(\underline{r}))$$

where λ is a numerical constant.

Object A is now removed, the photosensitive solid is treated so that its transmittance does not change further when light is incident upon it, and it is now illuminated by light reflected from object B, care being taken to maintain the spatial relationships between parts of the apparatus. At any point $\underline{r}$ in the solid (the hologram) assuming/ . .

assuming that the transmittance before the experiment was constant throughout the solid, the electric field amplitude is

$$\begin{aligned} G(\underline{r}) &= (t(\underline{r}) + \Delta t(\underline{r}))F_B(\underline{r}) \\ &= tF_B + \lambda(F_A+F_B)(F_A^*+F_B^*)F_B \\ &= tF_B + \lambda((F_AF_A^* + F_BF_B^*)F_B + F_BF_A^*F_B + F_AF_BF_B^*) \end{aligned}$$

(For brevity reference to the vector $\underline{r}$ is not made.)

With the assumption that the intensities I_A and I_B due to objects A and B are constant over the hologram, it is seen that

$$G = (\lambda(I_A+I_B) + t)F_B + I_BF_A + F_BF_A^*F_B$$

Thus as <u>both</u> wave fronts have been reconstructed, on looking through the hologram from the right in the direction of B and then in the direction of the original position of A (Figure 1B on page 26) <u>both</u> images are seen. That of B is brighter than that of A and the picture is marred by the presence of the wavefronts represented by the term $F_BF_A^*F_B$ in the above equation. As different parts of the hologram transmit these waves in different directions, the effect of this term can best be regarded as that of introducing noise into the picture.

Here the hologram is functioning as a memory. A record of A and B is stored and information about B is used as an address to retrieve the stored information about A (similarly A can be used to retrieve information about / . .

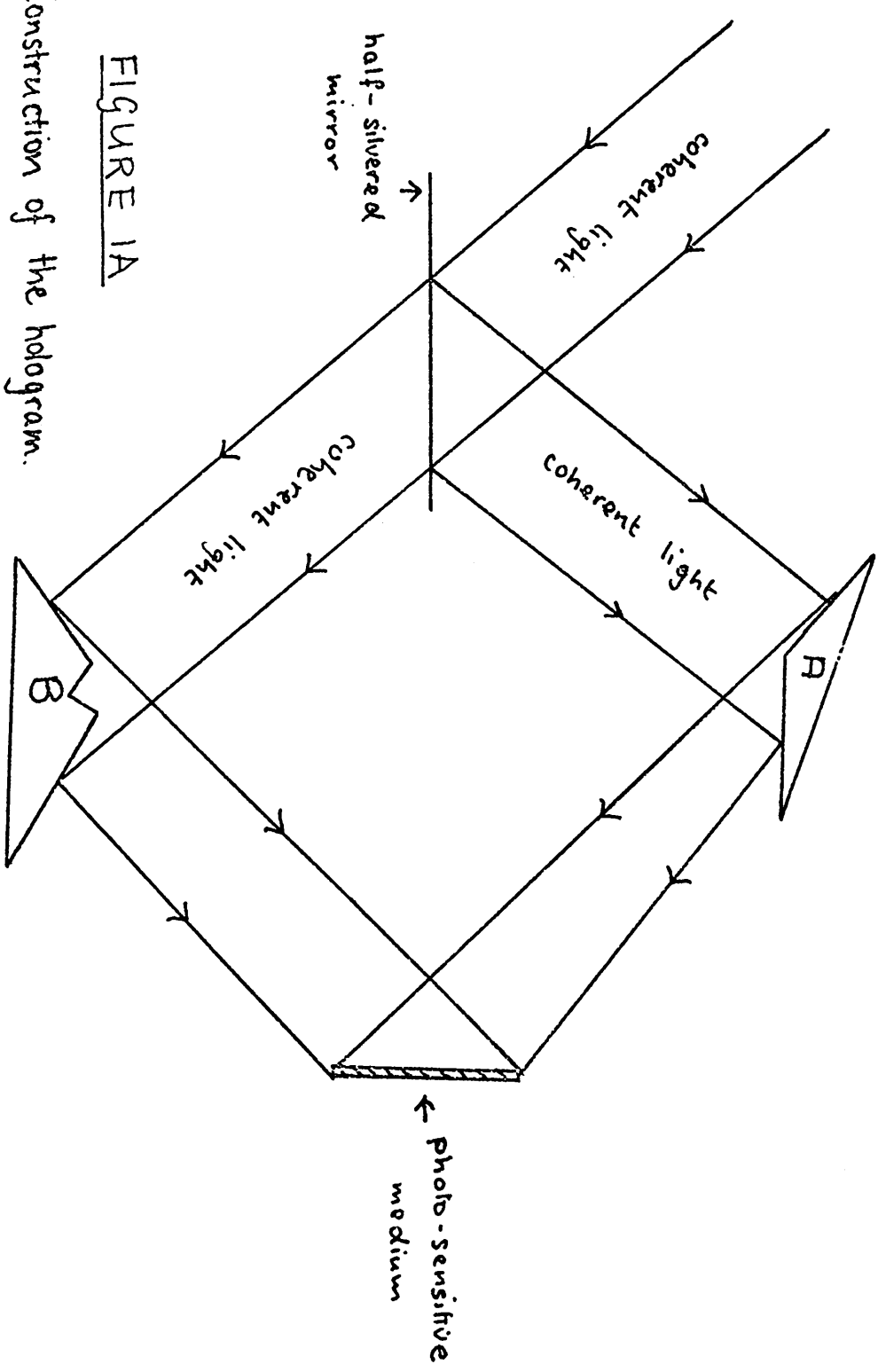

FIGURE 1A

Construction of the hologram.

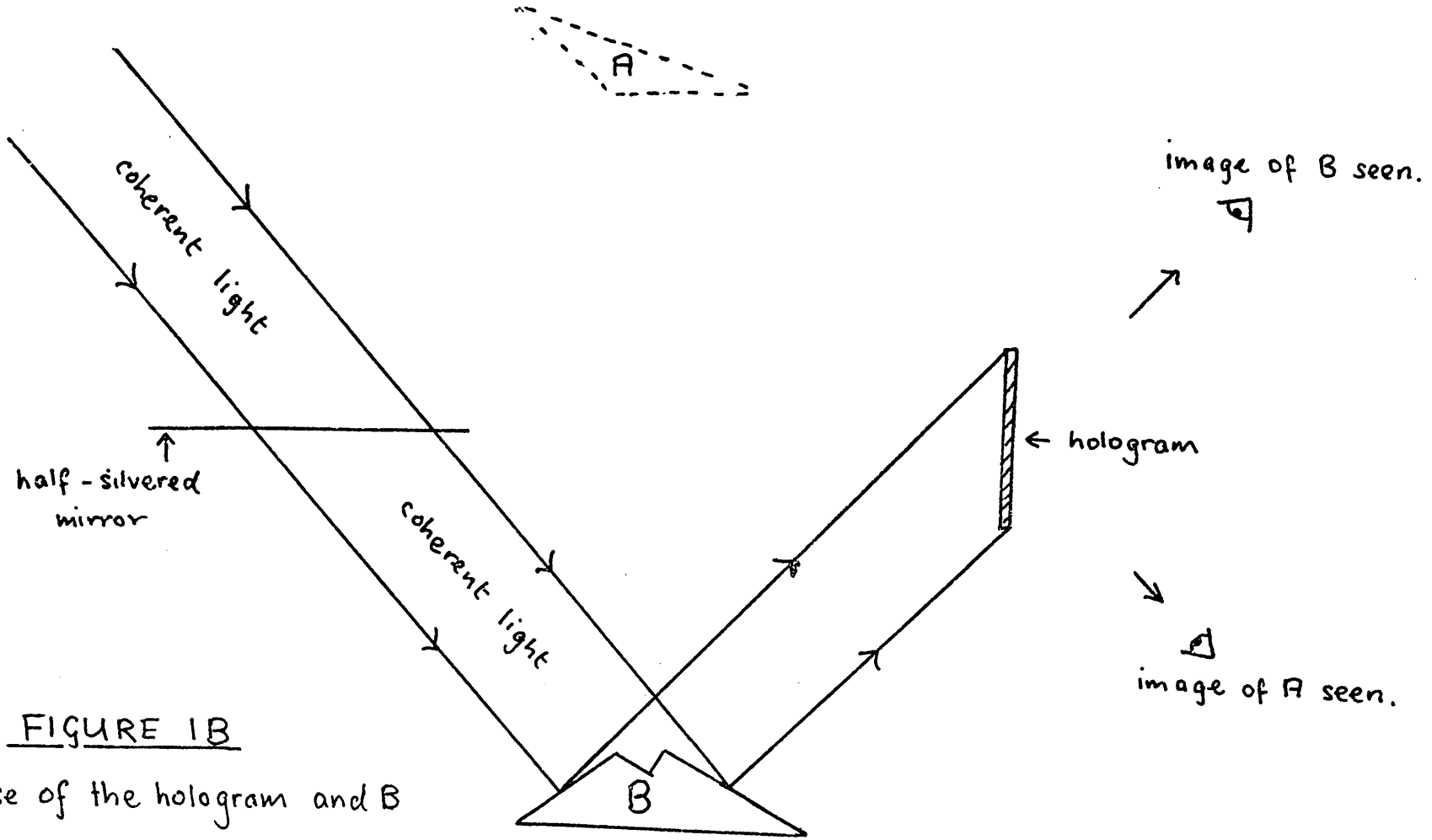

FIGURE 1B

Use of the hologram and B to produce an image of A.

about B). Each portion of the hologram contains information about each part of A and B from which it receives light. Furthermore, if it were possible to record information about more than one pair of objects by exposing the hologram in turn to different sets of interfering waves, then it would function as a distributed memory.

3.3 The Holophone

We will now turn to consider a content-addressable distributed memory which stores temporal signals. This is the Holophone which was invented by H.C.Longuet-Higgins (1968a,1968b). It comprises a large number of narrow band-width resonators arranged to cover uniformly a wide frequency range. They share a common input and a common output and each is connected to a variable gain amplifier which has as input a particular linear function of the instantaneous displacement and velocity of the appropriate resonator (Figure 2, page 28)

We suppose that each resonator, with displacement X , angular frequency w and damping constant ν , is given an acceleration by the incoming signal $F(t)$. The equation of motion is

$$\ddot{X} + 2\nu\dot{X} + (\nu^2 + w^2)X = F$$

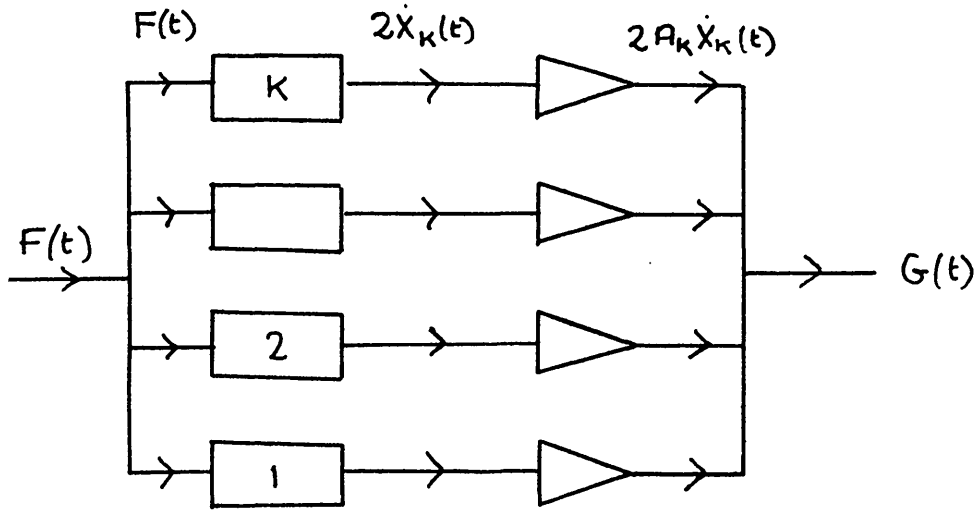

FIGURE 2 The Holophone

(The nomenclature refers to the ideal Holophone)

We will look at the performance of an ideal Holophone, that is we shall assume that the resonators are undamped. We realise that this assumption prevents our theory from being applied strictly to any practical situation, but we do this for the sake of simplicity of explanation and because we are only interested in the general principles underlying the behaviour of this system.

We thus rewrite the equation of motion of a resonator as

$$\ddot{X} + w^2X = F \quad \ldots\ldots\ldots\ldots 3.1$$

or / . .

$$\text{or } \left(\frac{d}{dt} + iw\right)\left(\frac{d}{dt} - iw\right)X = F$$

We now introduce the variables

$$Y = \dot{X} - iwX \quad \text{and} \quad Y^* = \dot{X} + iwX, \quad \ldots\ldots 3.2$$

so that we can write

$$\left(\frac{d}{dt} + iw\right)Y = \left(\frac{d}{dt} - iw\right)Y^* = F$$

$$\text{i.e.} \quad Y(t) = e^{-iwt} \int_{-\infty}^{t} F(\tau)e^{iw\tau}\, d\tau \quad \ldots\ldots 3.3$$

$$\text{and} \quad Y^*(t) = e^{+iwt} \int_{-\infty}^{t} F(\tau)e^{-iw\tau}\, d\tau \quad \ldots\ldots 3.4$$

For each resonator there is a device (represented by a triangle in Figure 2) which measures the function

$$Y(t) + Y^*(t) = 2\dot{X} \quad \ldots\ldots\ldots\ldots 3.5$$

This forms the input to the attached amplifier (For an ideal Holophone we do not also measure the instantaneous displacement of each resonator).

The final output from the Holophone is the sum of the outputs from all the amplifiers.

$$\text{i.e.} \quad G(t) = \sum_k A_k(Y_k(t) + Y_k^*(t))$$

where A_k is the gain of amplifier k.

We shall now assume that the resonant frequencies w_k (k=1,2,3,...) are so densely distributed over the given range that this sum can be replaced by an integral. Supposing that $w_k = k\mu$ and $w_{k+1} = (k+1)\mu$, where μ is very small,/ . .

small, then we write

$$G(t) = \frac{1}{\mu}\int_0^\infty A(w)(Y(w,t)+Y^*(w,t))dw \quad \ldots\ldots\ldots 3.6$$

where $A(w)$, $Y(w,t)$ and $Y^*(w,t)$ have replaced A_k, $Y_k(t)$, $Y_k^*(t)$.

The response function $M(\tau)$ of the Holophone is defined by the equation

$$G(t) = \int_0^\infty M(\tau)F(t-\tau)d\tau \quad \ldots\ldots\ldots\ldots 3.7$$

Let $F(\tau) = \delta(\tau)$. ($\delta(\tau)$ is the Dirac delta function)
Then, from 3.3, 3.4, and 3.7

$$Y(w,t) = e^{-iwt}, \quad Y^*(w,t) = e^{+iwt} \quad \text{for } t \geq 0.$$

and $G(t) = M(t)$,

so that from 3.6,

$$M(t) = G(t) = \frac{1}{\mu}\int_0^\infty A(w)(e^{-iwt} + e^{+iwt})dw \quad \ldots 3.8$$

$M(t)$ is only defined for $t \geq 0$.

Initially we set all the amplifier gains equal to the same real number A. In this case

$$M(t) = \frac{A}{\mu}\int_0^\infty (e^{-iwt} + e^{+iwt})dw$$

$$= \frac{2\pi A}{\mu}\delta(t) \quad \ldots\ldots\ldots\ldots 3.9$$

Thus initially the Holophone merely passes on the undistorted input amplified by an amount $\frac{2\pi A}{\mu}$.

<u>Storage</u>

Storage

In this process the signal F(t), which is over by the time t=0, is recorded by adjusting the gain of each amplifier by the amount

$$\Delta A(w) = \lambda \int_{-\infty}^{0} (Y(w,t) + Y^{*}(w,t))F(t)\,dt \quad . \quad . \quad .$$

where λ is the same for all resonators

This is possible for, as the signal F(t) is input, a record of $Y(w,t) + Y^{*}(w,t)$ has been kept. $\Delta A(w)$ can be expressed in another form

$$\Delta A(w) = 2\lambda \int_{-\infty}^{0} \dot{X}(w,t)(\ddot{X}(w,t)+w^{2}X(w,t))dt$$

(from 3.5 and 3.1)

$$= \lambda \int_{-\infty}^{0} \frac{\delta}{\delta t}(\dot{X}^{2}(w,t)+w^{2}X^{2}(w,t))dt$$

$$= \lambda \int_{-\infty}^{0} \frac{\delta}{\delta t}(Y(w,t)Y^{*}(w,t))dt \quad \text{(from 3.2)}$$

Finally, from 3.3 and 3.4,

$$\Delta A(w) = \lambda\, \phi(w)\phi^{*}(w)$$

where $\phi(w) = \int_{-\infty}^{0} F(t)e^{iwt}dt$

and $\phi^{*}(w)$ is the complex conjugate of $\phi(w)$

We see that the ideal Holophone, by virtue of the settings of its amplifier gains, records the power spectrum / . .

spectrum of the stored signal.

The gain of an amplifier is now

$$A(w) = A + \Delta A(w)$$
$$= A + \lambda \phi(w)\phi^{*}(w) \quad \ldots\ldots\ldots 3.10$$

It will be noted that a second signal H(t) can be stored by feeding it into the Holophone and turning up the amplifier gains as before. In general if R signals $F_r(t)$ (r=1,2, ...R) are recorded, then the amplifier gains have value

$$A(w) = A + \lambda \sum_{r=1}^{R} \phi_r(w)\phi_r^{*}(w) \quad \ldots\ldots 3.11$$

$$\text{where } \phi_r(w) = \int_{-\infty}^{0} F_r(t)e^{iwt}\,dt$$

Retrieval

We return, however, to the simpler situation where only one signal is stored. To use the Holophone to retrieve information, a signal $F'(t)$ which is a fragment of F(t), is now fed in to be used as the address or cue to retrieve the rest of F(t). Once again time is measured so that the instant t=0 occurs after the end of $F'(t)$ and for simplicity it will be assumed that each amplifier gain originally had the value $\frac{\mu}{2\pi}$ (see equation 3.9). After recording F(t), from 3.8 and 3.10, the response function of the Holophone is now

$$M'(\tau) = \delta(\tau) + \frac{\lambda}{\mu} \int_{-\infty}^{+\infty} \phi(w)\phi^{*}(w)e^{-iwt}\,dw$$

(We can extend the limits of integration if we let w take positive and negative values).

Consequently, using 3.7, the output $G'(t)$ is now not proportional to the input but is of the form

$$G'(t) = F'(t) + \Delta G(t) \quad \ldots\ldots\ldots 3.12$$

where

$$\Delta G(t) = \frac{\lambda}{\mu} \int_{-\infty}^{+\infty} dw \int_{0}^{\infty} d\tau \; \phi(w)\phi^*(w)e^{-iw\tau} F'(t-\tau)$$

Writing $u = t-\tau$, then we finally express this extra output $\Delta G(t)$, which we shall later show to resemble the stored signal $F(t)$, as

$$\Delta G(t) = \frac{\lambda}{\mu} \int_{-\infty}^{+\infty} dw \int_{0}^{-\infty} du \; \phi(w)\phi^*(w)e^{-iw(t-u)}F'(u)$$

$$= \frac{\lambda}{\mu} \int_{-\infty}^{+\infty} dw \int_{-\infty}^{0} du \; e^{iwt}\phi(w)\phi^*(w)e^{-iwu}F'(u)$$

$$\ldots 3.13$$

$$\text{where } \phi(w) = \int_{-\infty}^{0} F(t)e^{iwt}dt$$

3.4 Discrete Formulation of the Holophone

A convenient way of expressing the mathematics of this ideal Holophone is obtained by considering its discrete analogue (It is, of course, a discrete device anyway, in that its resonators are not spaced continuously over an infinite range of frequency).

R signals, each of which lasts N units of time, are stored./ . .

stored. The N successive real values of the amplitudes of one of these signals, for example signal n, are placed in one column $F^{(n)}$ of an NxR matrix F_{ij} (i=0,1,2,..N-1. j=1,2,..R). The discrete Fourier Transform of $F^{(n)}$ is

$$\phi^{(n)} = MF^{(n)}$$

i.e. $$\phi_{jn} = \sum_{k=0}^{N-1} M_{jk}F_{kn}, \text{ where } M_{jk}=\frac{1}{N}\exp(\frac{2\pi i}{N} jk)$$

By analogy to equation 3.11, the change in the gain of the k^{th} amplifier after storing R signals is

$$\Delta A_k = \lambda \sum_{r=1}^{R} \phi_{kr}\phi^{*}_{kr},$$

and the values of these gains are contained in a diagonal matrix D with components

$$D_{jk} = \lambda\delta_{jk} \sum_{r=1}^{R} \phi_{jr}\phi^{*}_{kr} \quad \text{where } \delta_{jk} \text{ is the Krönecker delta.}$$

When a signal $F'^{(n)}$, which is a portion of $F^{(n)}$ and lasts N' units of time, is input, the discrete counterpart of equation 3.12 is $G'^{(n)} = F'^{(n)} + \Delta G^{(n)}$,

with $$\Delta G^{(n)} = \gamma M^{-1} DMF'^{(n)} \quad \ldots\ldots\ldots \quad 3.14$$

where D is defined as above, M^{-1} is the inverse Fourier Transform matrix and γ is a numerical constant. The input to the Holophone, represented by the column vector $F'^{(n)}$, has components $F'_{in} = F_{i+p,n}$ for i=0,1,2,..N'-1

and $F'_{in} = 0$ otherwise $(P \geqslant 0)$

The / . .

The j^{th} component of $\Delta_G^{(n)}$ is

$$\Delta G_{jn} = \lambda\gamma \sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1} M^{-1}_{jk} D_{kl} M_{lm} F'_{mn}$$

$$= \sum_{r=1}^{R}\sum_{k,l,m=0}^{N-1} M^{-1}_{jk}\delta_{kl}\sum_{s=0}^{N-1} M_{ks}F_{sr}\sum_{t=0}^{N-1} M^{*}_{lt}F^{*}_{tr}M_{lm}F'_{mn}$$

$$= \frac{\lambda\gamma}{N^4}\sum_{r=1}^{R}\sum_{m,s,t,l=0}^{N-1} \exp\left(\frac{2\pi i}{N}(s+m-t-j)l\right)F_{sr}F_{tr}F'_{mn}$$

(note that F has real components and that some summation signs are omitted)

$$= \frac{\lambda\gamma}{N^3}\sum_{r=1}^{R}\sum_{m,s,t=0}^{N-1} \delta_{s,t+j-m}F_{sr}F_{tr}F'_{mn} \quad \ldots \quad 3.15$$

$$= \frac{\lambda\gamma}{N^3}\sum_{r=1}^{R}\sum_{m,s=0}^{N-1} F_{sr}F_{s+m-j,r}F'_{mn} \quad \ldots \quad 3.16$$

In equation 3.15 consider the quantity

$$\Delta_r = \sum_{m,s,t=0}^{N-1} \delta_{s,t+j-m}F_{sr}F_{tr}F'_{mn}$$

This can be written in several forms depending on which variable is eliminated in simplifying it.

(i) Substituting for s we have

$$\Delta_r = \sum_{m=0}^{N-1}\sum_{t=0}^{N-1} F_{t+j-m,r}F_{tr}F'_{mn} = \sum_{m=0}^{N-1} A_{j-m,r}F'_{mn} \quad \ldots \quad 3.17$$

where $A_{j-m,r} = \sum_{t=0}^{N-1} F_{t+j-m,r}F_{tr}$.

This is the autocorrelation function $A^{(r)}$ of $F^{(r)}$.

(ii) / . .

(ii) Alternatively, by eliminating m we write

$$\Delta_r = \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} F_{sr} F_{tr} F'_{t+j-s,n} = \sum_{s=0}^{N-1} C_{j-s,n,r} F_{sr} \quad . \quad . \quad 3.18$$

where $C_{j-s,n,r} = \sum_{t=0}^{N-1} F_{tr} F'_{t+j-s,n}$

This is the cross-correlation function $C^{(n,r)}$ of $F^{(r)}$ with $F'^{(n)}$.

Writing 3.15 in the form

$$\Delta G_{jn} = \frac{\lambda\gamma}{N^3} \left(\sum_{m,s,t=0}^{N-1} \delta_{s,t+j-m} F_{sn} F_{tn} F'_{mn} + \sum_{m,s,t=0}^{N-1} \sum_{\substack{r=1 \\ r \neq n}}^{R} \delta_{s,t+j-m} F_{sr} F_{tr} F'_{mn} \right) ,$$

Then, using 3.17 and 3.18, this becomes

$$\Delta G_{jn} = \frac{\lambda\gamma}{N^3} \left(\sum_{s=0}^{N-1} C_{j-s,n,r} F_{sn} + \sum_{m=0}^{N-1} \sum_{\substack{r=1 \\ r \neq n}}^{R} A_{j-m,r} F'_{mn} \right) \quad . \quad . \quad 3.19$$

If the stored signals are random (i.e. each component is chosen independently) then

$$A_{j-m,r} = \sum_{k=0}^{N-1} F_{j+k-m,r} F_{kr} \approx N \delta_{jm}$$

Furthermore, if the portion $F'^{(n)}$ of $F^{(n)}$ resembles $F^{(n)}$ sufficiently closely, then

$$C_{j-s,n,n} = \sum_{t=0}^{N-1} F_{tn} F'_{t+j-s,n}$$

$$= / \ . \ .$$

$$= \sum_{t} F_{tn} F_{t+j-s+p,n'} \quad \text{where t runs over N' numbers only}$$

$$\simeq N' \delta_{j,s-p}$$

Equation 3.19 may then be written as

$$G_{jn} \simeq \frac{\lambda\gamma}{N^3}\left(\sum_{s=0}^{N-1} N' \delta_{j,s-p} F_{sn} + N(R-1)\sum_{n=0}^{N-1} \delta_{jm} F'_{mn}\right)$$

$$= \frac{\lambda\gamma}{N^3}\left(N' F_{j+p,n} + N(R-1)F'_{jn}\right)$$

The output of the Holophone is consequently

$$G'_{jn} = F'_{jn} + \Delta G_{jn} \simeq F'_{jn}\left(1 + (R-1)\frac{\lambda\gamma}{N^2}\right) + \frac{\lambda\gamma N'}{N^3} F_{j+p,n}$$

Now $F'_{jn} = F_{j+p,n}$ for $j=0,1,2,..N'-1$

and $F'_{jn} = 0$ otherwise.

Thus for times after the end of the cue $F'^{(n)}$ (for values of j greater than $N'-1$), the Holophone continues to respond, the output being a noisy version of the signal $F^{(n)}$ originally stored and bearing the correct temporal relationship to the cue. The output is

$$G'_{jn} \simeq \left(1 + \frac{\lambda\gamma}{N^3}\left(N(R-1)+N'\right)\right) F_{j+p,n} \quad \text{for } j=0,1,2,..N'-1$$

and $G'_{jn} \simeq \frac{\lambda\gamma N'}{N^3} F_{j+p,n}$ for $j=N', N'+1, .. N-1$

The Holophone's output was termed noisy owing to the incorrect assumptions we have made about particular properties of the cross-correlation and auto-correlation functions of random / . .

random signals. Before a more detailed statement about this is made, the Holophone will be put in its precise holographic context by describing an analogous holographic model.

3.5 Holographic analogue of the Holophone

Coherent laser light illuminates a transparency T of a simple plane figure (for example a triangle) placed in the focal plane P_0 of lens L_1 (Figure 3 on page 39). An inverted image of the triangle is seen on the screen S which is in the focal plane P_2 of Lens L_2. Lens L_1 and L_2 are so arranged that their two other focal planes (one from each lens) coincide (Longuet-Higgins et al.,1970).

A photographic plate is now placed in the common plane P_1 and is exposed to the light passing through the transparency. It is then removed, a positive transparency, the hologram H, is made of it and put in its place. Part of the transparency is now masked. An inverted, although distorted, image of the whole of the original scene is still seen on the screen S. Translating this situation into the terms of the more general holographic set up described in Chapter 2, the unmasked section of the transparency is object B, the masked section object A. As before, object B is used as the address to retrieve a record of object A from the memory.

Analysis is possible if the following fact is used. When / . .

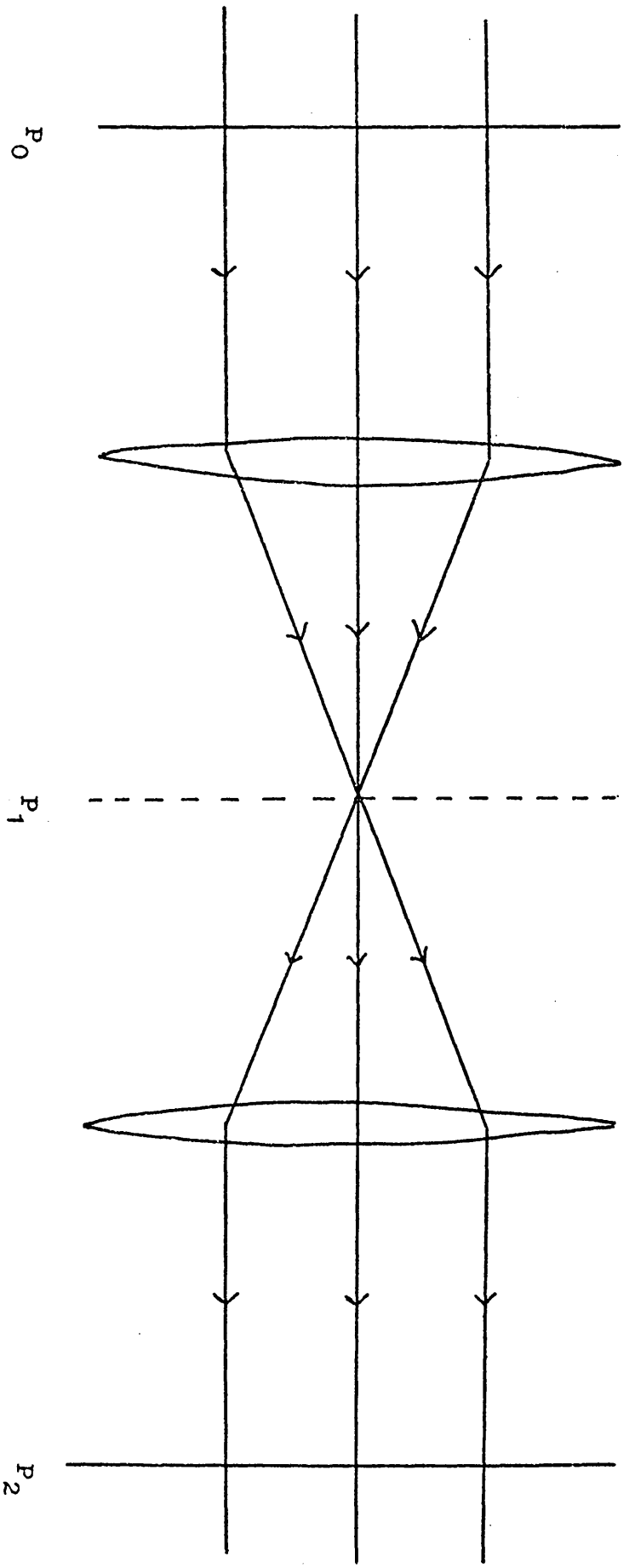

FIGURE 3 A holographic analogue of the Holophone.

When coherent light passes through both focal planes of a perfect lens then the amplitude of the electric field in one plane is the Fourier Transform of the amplitude in the other. Denoting the field at any point (x_i, y_i) in the plane P_i (the coordinates x_i and y_i being measured with respect to an origin in the plane) by the symbol E_i, then, with the hologram absent, the field amplitude at a point in plane P_1 is

$$E_1(x_1, y_1) = \iint E_0(x_0, y_0)\, e^{\frac{2\pi i}{\lambda f}(x_0 x_1 \cos\theta_o + y_0 y_1 \cos\phi_o)} dx_0 dy_0$$

Similarly, for plane P_2

$$E_2(x_2, y_2) = \iint E_1(x_1, y_1)\, e^{\frac{2\pi i}{\lambda f}(x_1 x_2 \cos\theta_1 + y_1 y_2 \cos\phi_1)} dx_1 dy_1$$

where, for example, $\cos\theta_o$ and $\cos\phi_o$ are the directional cosines of light from plane P_o incident at the point (x_1, y_1). These can both be set equal to 1, provided that $a^2 \ll f^2$, where 2a is the length of one side of the transparency (assumed square).

Therefore, combining these equations

$$E_2(x_2, y_2) = \iiint E_o(x_0, y_0) e^{\frac{2\pi i}{\lambda f} x_1 (x_0 + x_2)} e^{\frac{2\pi i}{\lambda f} y_1 (y_0 + y_2)} dx_1 dy_1 dx_0 dy_0$$

$$\propto \iint E_0(x_0, y_0)\ \delta(x_0 + x_2)\ \delta(y_0 + y_2)\, dx_0 dy_0$$

$$= E_0(-x_2, -y_2)$$

Thus, in the absence of the hologram, an inverted image/ . .

image of the triangle is seen on the screen.

Now, to consider what happens with the hologram in place, the functions F_0 and G_0 will be introduced. $F_0(x_0,y_0)$ is the field at a point (x_0,y_0) in plane P_0 owing to light transmitted from a fragment F of the transparency.

Assuming the change in the hologram transmittance t to be linear with respect to the intensity of light falling upon it

$$\text{i.e. } \Delta t(x_1,y_1) = \lambda E_1(x_1,y_1)E_1^*(x_1,y_1),$$

then, when light from the fragment F strikes the hologram, the field $EE_1(x_1,y_1)$ at a point (x_1,y_1) in plane P_1, equals

$$\iint (t+ \lambda E(x_1,y_1)E_1^*(x_1,y_1))F_0(x_0,y_0)e^{\frac{2\pi i}{\lambda f}(x_1x_0+y_1y_0)}dx_0dy_0$$

and in the plane P_2 the field $EE_2(x_2,y_2)$ is

$$\iint e^{\frac{2\pi i}{\lambda f}(x_1x_2+y_1y_2)}EE_1(x_1,y_1)dx_1dy_1$$

$$= KtF_0(-x_2,-y_1)$$

$$+ \lambda \iint e^{\frac{2\pi i}{\lambda f}(x_1x_2+y_1y_2)}E_1(x_1,y_1)E_1^*(x_1y_1)F_1(x_1,y_1)dx_1dy_1$$

where $F_1(x_1y_1) = \iint e^{\frac{2\pi i}{\lambda f}(x_0x_1+y_0y_1)}F_0(x_0y_0)dx_0dy_0$

and K is a constant.

We consider the case of a one-dimensional hologram. For this, the electric field at a point on the line P_2 is

$$EE_2(x_2) = KtF_0(-x_2)+ \lambda \int dx_1 \int dx_0 e^{\frac{2\pi i}{\lambda f}x_1x_2}E_1(x_1)E_1^*(x_1)e^{\frac{2\pi i}{\lambda f}x_1x_0}F_0(x_0)$$

where $E_1(x_1) = \int E_0(x_0) e^{\frac{2\pi i}{\lambda f} x_0 x_1} dx_0.$

We compare this with the output from the Holophone - namely

$$G'(t) = F'(t) + \frac{\lambda}{\mu} \int_{-\infty}^{+\infty} dw \int_{-\infty}^{0} du\ e^{-iwt}\ \phi(w)\phi^*(w) e^{iwu} F'(u)$$

(equations 3.12 and 3.13)

It can be seen that a one-dimensional hologram of the type described performs the same job in the spatial domain as an ideal Holophone does in the time domain (apart from trivial differences such as image inversion in the holographic case). In particular, the grains of the holographic plate correspond to the ideal Holophone's amplifiers. Each amplifier or grain contains information about the signal to be stored and if it were possible to subject the hologram to multiple exposures then, like the Holophone, it would perform as a distributed memory.

3.6 Storage and retrieval of random binary signals by the Holophone

The performance of a Holophone (or for that matter a holographic model of the type just described) will be discussed in the particular case of the storage of random binary signals - that is the components of the vectors which represent the signals may take the value +1 or -1 with equal probability. Cyclic boundary conditions are imposed, each component of a vector being associated with a vector of an N-sided polygon.

Analysis Defining for simplicity $\Delta G'_{jn} = \frac{N^3}{\lambda\gamma} \Delta G_{jn}$, by substituting / . .

substituting for $A^{(n)}$ and $C^{(n,r)}$, equation 3.19 will be written as

$$\Delta G'_{jn} = \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} F_{tn} F'_{t+j-s,n} F_{sn} + \sum_{\substack{r=1 \\ r\neq n}}^{R} \sum_{m=0}^{N-1} \sum_{k=0}^{N-1} F_{j+k-m,r} F_{kr} F'_{mn} \quad \ldots 3.20$$

$$= \sum_{s} \sum_{t} F_{tn} F_{t+j+p-s,n} F_{sn} + \sum_{r\neq n}^{R} \sum_{m} \sum_{k} F_{j+k-m,r} F_{kr} F_{m+p,n} \quad \ldots 3.21$$

Values of j relating to times after the end of the cue will be considered so that in the first sum of equation 3.21 the quantity t+j+p-s may only have N' different values. The same remark applies to the suffix m+p in the second sum. In particular, since $F'_{jn}=0$ (i.e. F_{j+p} is not one of the components of the cue), $t+j+p-s\neq j+p$ i.e. $t\neq s$. Also $m+p\neq j+p$ i.e. $m\neq j$.

The second term of equation 3.21 is a sum of (R-1)NN' numbers, each of which has value +1 or -1 equiprobably, independent of the value of other items, except where $j+k-m=k$ i.e. $m=j$. This never arises.

The first term contributes NN' numbers of value +1 or -1 which are again independent and equally likely to have the value +1 as -1 unless values of suffices coincide. There are three possibilities.

(i) $t=t+j+p-s$. Picking out the relevant terms from 3.21, the contribution to $\Delta G'_{jn}$ is

$$\sum_{t} / \ldots$$

$\sum_t F_t F_t F_{j+p,n} = N' F_{j+p,n}$ (t=t+j+p-s in only allowed N' different values)

(ii) t = s. From 3.20, contributions to $\Delta' G_{jn}$ are coefficients of F'_{jn} which takes the value 0.

(iii) t+j+p-s=s. The contribution to $\Delta G'_{jn}$ is

$$\sum_s F_{2s-j-p,n} F_{t+j+p-s,n} F_{t+p+j-s,n} = \sum_s F_{2s-j-p,n}$$

This is a sum of independent terms.

There are no values of s and t for which more than one of these conditions are satisfied.

Thus $\Delta G'_{jn}$ is made up of NN'(R-1) + (N-1)N' terms each of mean 0 and variance 1 and N' terms each of mean F_{jn} and variance 0. Summing over the ensemble of the components of the output signal, $\Delta G'_{jn}$ has mean $N' f_{jn}$ and variance N'(NR-1).

The fidelity of reproduction of the retrieved signal can be expressed in terms of the signal to noise ratio S/N defined as the ratio of the square of the mean of a component of output signal to its variance. We thus have

$$S/N = \frac{N'^2}{N'(NR-1)} \simeq \frac{N'}{NR} \quad (\text{for } NR \gg 1) \quad \ldots \quad 3.22$$

which is in fact equal to the ratio of the length of the cue to the total length of the remaining signals in store. This result has been checked by computer simulation using ALGOL as the programming language.

Computer Simulation

The/ . .

The following procedure was adopted:

1) N components of R input signals were generated with the aid of a pseudo-random number generator.

2) The first N' of each of these were used as cues and the N-N' components of the R recalled signals $G^{(n)}$ (n=1,2,..R) were calculated, using 3.14.

3) Of these components approximately half arose from components of $F^{(n)}$ of value +1 and for this subset the signal to noise ratio was calculated from the formula

$$+ve\ S/N = (G_{av})^2/((G^2)_{av}-(G_{av})^2)$$

4) The same was done for the components arising from signal components of value -1 and the two results were averaged to give a figure for the overall signal to noise ratio for the recalled parts of the R stored signals.

Three situations were considered; those when 1, 5 and 10 signals were stored. In each case cues of different lengths were used in recall.

With R=1, N was set at 401 and recall from cues of length N' = 50, 100, 150, 200, 250 and 350 was considered.

With R=5 and R=10, N was set at 201 and N' varied from a value of 25 to 175 in steps of 25.

For a given set of values of R, N and N', different subsets of signals drawn from the same pseudo-random/

gave
random ensemble/ different values of the signal to noise ratio. Thus for each triad of values of R,N and N', 5 pseudo-randomly chosen sets of signals were recorded and retrieved and the mean and standard deviation of the Signal to Noise Ratio were calculated for this population of 5.

These results are portrayed in graphical form in Figure 4 (see pages 47, 48 and 49). It would seem that the theory is adequate in producing an approximate expression for the variation of the signal to noise ratio of the Holophone with cue length although it is perhaps optimistic when the cue is almost as long as the stored message. It was not thought worthwhile for the program to be rewritten in machine code employing Fast Fourier Transforms so that the feasible values of N and R could be increased, thereby producing more reliable results.

For the signals to be retrieved reasonably accurately the signal to noise ratio must be large compared to 1. This condition is never fulfilled and we must conclude that the Holophone does not perform this task very well.

3.7 The Off-Diagonal Holophone

The Holophone is, however, just one example of the family of linear distributed memory models which are described by the matrix equation

$$G'^{(n)} = F'^{(n)} + \Delta G'^{(n)}$$

where $\Delta G'^{(n)} = M^{-1} D M F'^{(n)}$ . . . . . .(3.14)

Each/ . . (compare 3.14 with equation 1.1 of section 1.6)

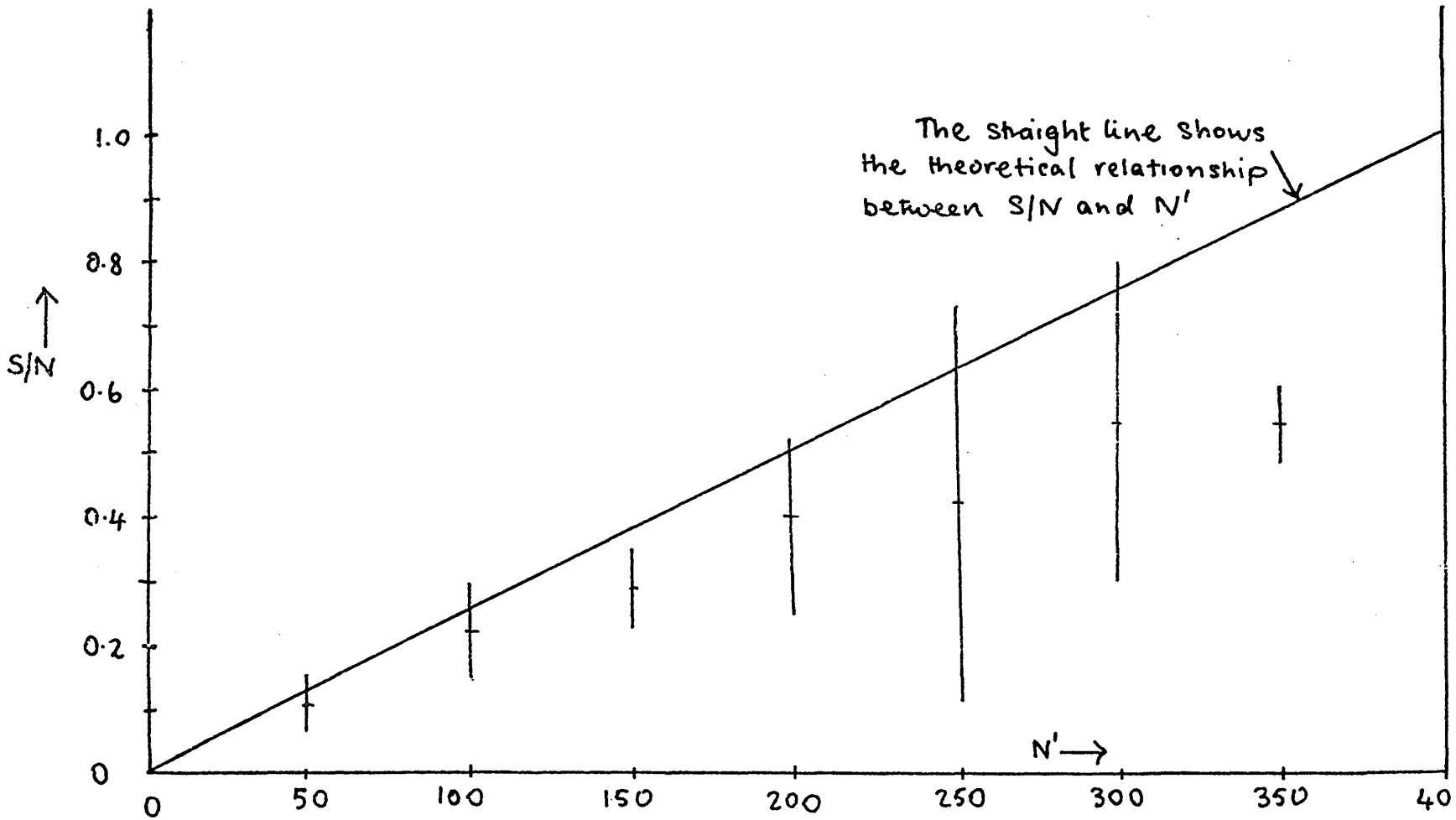

FIGURE 4A. S/N plotted as a function of N'. N = 401, R = 1. Means and standard deviations of S/N are shown.

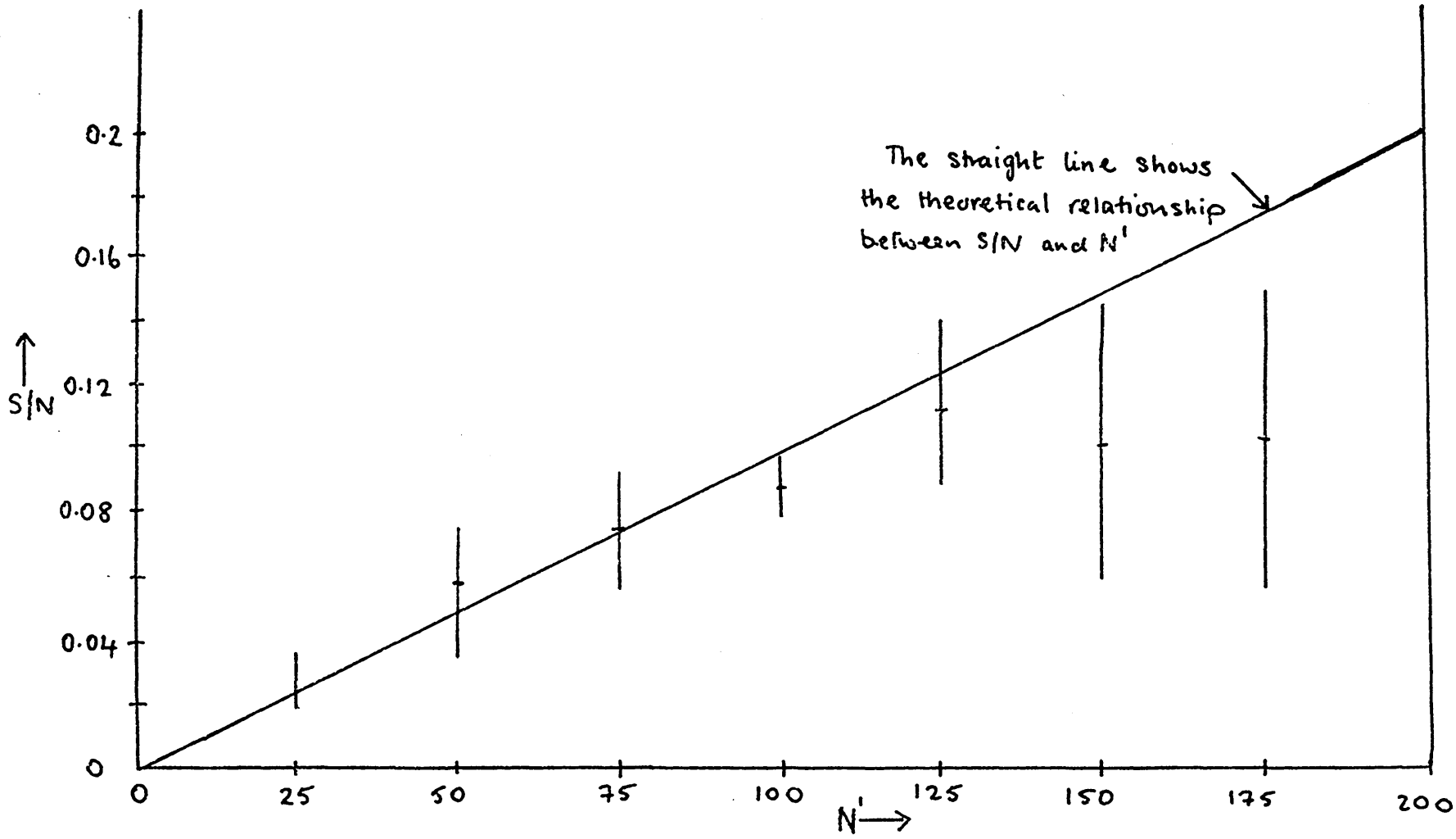

FIGURE 4B. S/N plotted as a function of N'. N=201, R=5.
Means and standard deviations of S/N are shown.

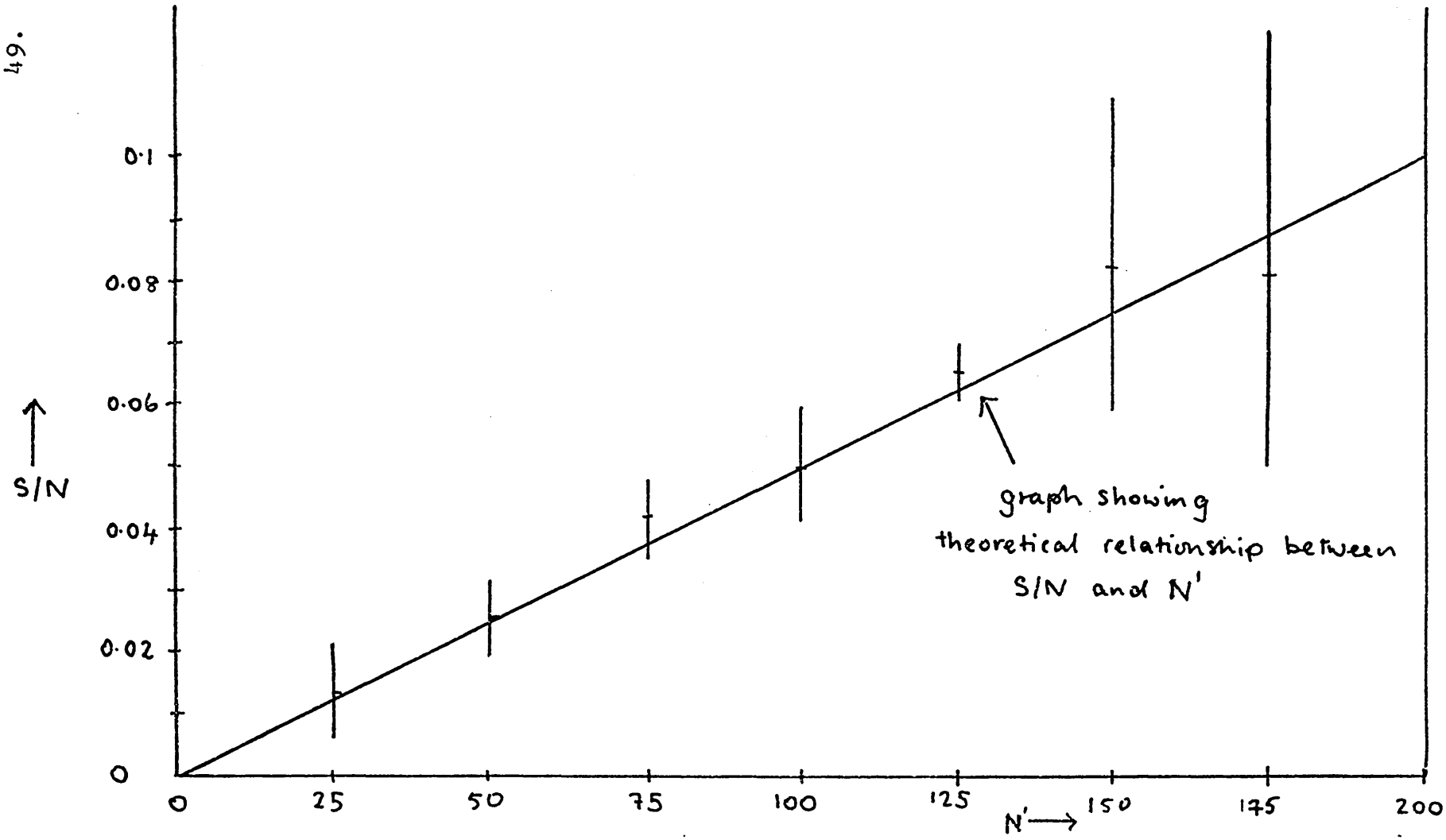

FIGURE 4C. N=201, R=10. Means and standard deviations of S/N are shown.

Each model is defined by a particular choice of M and the memory matrix D, the values of whose components are some function of the stored signals F, and equation 3.14 shows the output $G'^{(n)}$ from the device when a portion $F'^{(n)}$ of a stored signal is input. There is no reason to suppose that the form of D and M chosen for the Holophone produces the most efficient and reliable memory system.

There is a model amenable to analysis which is almost identical to the Holophone, except that the storage matrix is no longer diagonal but is given by

$$D_{pq} = \sum_{r=1}^{R} \phi_{pr}\phi^*_{qr}$$

$$= \sum_{r=1}^{R} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} M_{ps}F_{sr}M^*_{qt}F^*_{tr} \quad \ldots\ldots\ldots 3.23$$

We call this model the Off-diagonal Holophone. From 3.14 and 3.23 we can then write

$$\Delta G'_{jn} = \sum_{r=1}^{R} \sum_{p,s,t,q,w} M^{-1}_{jp} M_{ps}F_{sr}M^*_{qt}F^*_{tr}M_{qw}F'_{wn}$$

$$= \frac{1}{N^4}\sum_{r} \sum_{p,s,t,q,w} e^{\frac{2\pi i}{N}(-jp+ps-qt+qw)} F_{sr}F_{tr}F'_{wn}$$

$$= \frac{1}{N^2}\sum_{r} \sum_{s,t,w} \delta_{sj}\delta_{tw}F_{sr}F_{tr}F'_{wn}$$

$$= \frac{1}{N^2}\sum_{r} \sum_{t} F_{jr}F_{tr}F'_{tn} \quad \ldots\ldots\ldots\ldots 3.23A$$

$$= / \ .\ .$$

$$= \frac{1}{N^2}\sum_r \sum_t F_{jr}F_{tr}F_{t+p,n} \quad \ldots\ldots 3.24$$

This is a double sum of scalar products over rows and columns of the matrix F.

3.8 Storage and Retrieval of random binary signals by the Off-diagonal Holophone

Considering once again the retrieval of a random binary signal of length N from a cue of length N' then 3.23 is rewritten (dropping the constant $1/N^2$) as

$$\Delta' G_{jn} = \sum_s F_{jn}F_{sn}F'_{sn} + \sum_{r\neq n}\sum_s F_{jr}F_{sr}F'_{sn}$$

Values of j appropriate to moments of time after the end of the cue will be considered ($F'_{jn}=0$). Noting that $F'_{jn} = F_{j+p,n}$ ($p\geq 0$), the two situations p=0 and p>0 will be dealt with separately.

(i) p=0 i.e. $F'_{jn} = F_{jn}$ for j=0,1,2,.....N'-1

$$\text{Then } \Delta G'_{jn} = \sum_{s=0}^{N-1} F_{jn}F_{sn}F'_{sn} + \sum_{r\neq n}^{R}\sum_{s=0}^{N-1} F_{jr}F_{sr}F'_{sn}$$

$$= N'F_{jn} + \sum_{r\neq n}^{R}\sum_{s=0}^{N-1} F_{jr}F_{sr}F'_{sn}$$

The second term of this expression is a sum (R-1)N' terms which are all equally likely to have the value +1 as -1 unless s=j.

Setting s=j, $F_{jr}F_{sr}F'_{sn} = F_{jr}F_{jr}F'_{jn} = 0$ since $F'_{jn}=0$, Consequently $\Delta G'_{jn}$ has mean $N'F_{jn}$, variance (R-1)N' and the/ . .

the signal to noise ratio is

$$S/N = \frac{N'}{R-1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad 3.25$$

(ii) $p \neq 0$ $\qquad \Delta G'_{jn} = \sum_{r=1}^{R} \sum_{s=0}^{N'-1} F_{jr} F_{sr} F_{s+p,n}$

which is a sum of N'R independent terms of magnitude +1 or -1. The signal to noise ratio is zero.

We conclude that this particular model is extremely good at recalling messages provided that it is known from which part of the message the cue was taken (so that p is set equal to zero). In fact, subject to this condition, if only one message is stored then, whatever the length of the cue, retrieval is perfect (the signal to noise ratio is infinite).

3.9 Summary

A particular model of distributed memory called the Holophone has been discussed. It was found that it did not perform a particular storage and retrieval task very well. A related model which was discussed only in terms of the equations describing its functioning was found to perform this task better. Nothing has been said about a possible neuro-physiological implementation of the Holophone and at this point all that will be said is that like all holographic models it does place severe demands on the nervous system. For example, it is necessary that there be a large number of oscillating circuits each of which remain/ . .

remain sharply tuned to a particular frequency.

However, since the Holophone and the Off-diagonal Holophone both perform simple tasks, the former calculating cross-correlations and auto-correlations and the latter scalar products then, in the hope that structurally simpler models might have more neurophysiological relevance, the properties of less complex physical representations of both models have been considered. These models will be introduced in the next chapter.

CHAPTER 4

Non-holographic models

The Correlograph and the Associative Net

4.1 Introduction

In a search for simpler representations of holographic models of memory we shall take up the observation of Gabor (Gabor 1968a, 1968b, 1969) that a system which computes cross-correlations (or convolutions) can mimic the performance of a Fourier holograph.

4.2 A spatial linear analogue of the holophone

Such a system is shown in Figure 5 (see page 55). Two transparencies A and B are illuminated by means of a diffuse light source D. Light transmitted through B is focussed onto a plate C by means of the lens L. A and C are separated by a distance equal to twice the focal length of the lens and B is placed half way between them. We will assume that the transmittance at any point on plate C varies in proportion to the intensity of light falling on it. After exposure, plate C is developed, converted into a positive transparency, now called a Linear Correlogram, and replaced. The retrieval process comprises illuminating the Linear Correlogram by the diffuse source, shifting the lens to the other side of plate B and looking at the pattern of light falling on the screen S which has replaced plate A. By repeating the storage procedure with other pairs of patterns and using the same plate C it can be/ . .

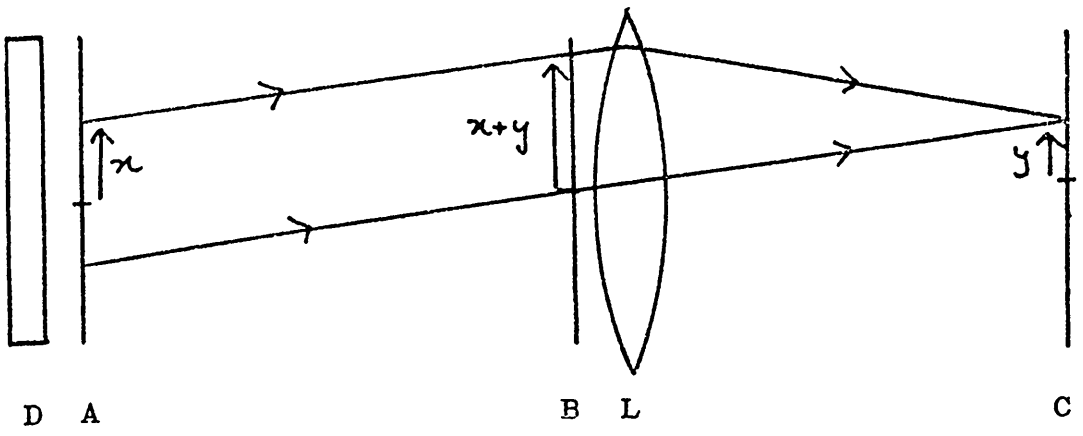

FIGURE 5A

Construction of the Linear Correlogram.

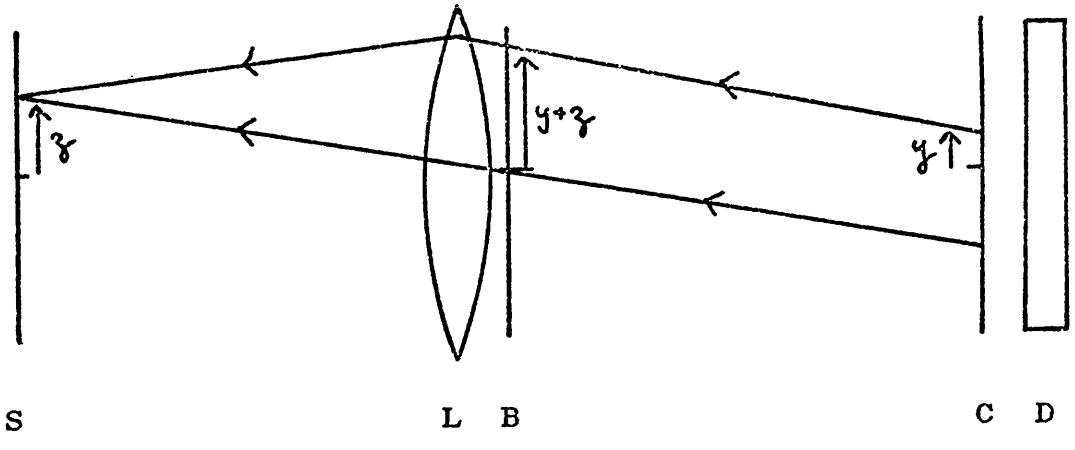

FIGURE 5B

Retrieval of pattern A using

the Linear Correlogram and pattern B.

(The distances x, y and z refer to the one-dimensional case)

be made to function as a distributed memory store. The complete system is called a Linear Correlograph.

For simplicity a one-dimensional Linear Correlograph will be considered.

The Linear Correlogram stores cross-correlations. Let A(x) and B(x+y) refer to the intensity of light transmitted through transparencies A and B at points x and x+y which contribute to the intensity C(y) at point y on C. x, x+y and y are measured along parallel lines drawn in planes A, B and C respectively. Each line cuts the common principal optic axis of the lenses at right angles.

The transmittance at point y in C is modified by an amount proportional to

$$C(y) = \int A(x)\ B(x+y)dx.$$

Retrieval also involves cross-correlation. The intensity at a point z on the viewing screen S is (disregarding numerical constants)

$$S(z) = \int C(y)B(y+z)dy$$

$$= \iint A(x)B(x+y)B(y+z)dydz.$$

The situation analogous to that with which the Holophone has to cope is that where auto-correlations are stored (A is identical to B) and a fragment A' of a stored pattern is used to recall the rest of it.

In that case $S(z) = \int C(y)A'(y+z)dy$

where $C(y) = \int A(x)A(x+y)dx.$

Writing u=y+z / . .

Writing u=y+z then

$$S(z) = \int C(u-z)A'(u)du \quad \ldots\ldots 4.1$$

With nomenclature as for equation 4.1, the continuous form of equation 3.16 describing the relevant part of the output of the Holophone is

$$S_H(z) = \iint A(x)A(x+u-z)A'(u)dudv$$
$$= \int C(u-z)A'(u)du$$
$$= S(z) \text{ (disregarding constant multiplying factors)}$$

Thus a particular experimental arrangement of the Linear Correlogram does the same job as the Holophone without the need of the latter's complex circuitry. Moreover, the Linear Correlogram is a more general associative memory. Information about pairs of patterns A and B is stored and pattern B, which is not necessarily part of A, can be used to retrieve pattern A. (It should be mentioned that by inverting plate C, pattern A can be used to retrieve pattern B.).

4.3 The Correlograph

4.3.1 Description

The Linear Correlograph and the Holophone are linear devices according to our definition made in section 1.6. We will now discuss a related non-linear device (which we shall later show to be non-linear in the sense that this word has been used in Chapter 1) which is amenable to detailed analysis. We will call this the Correlograph.

A/ . .

A and B are black cards on which are punched patterns of pinholes. These, when illuminated by the light source D produce a pattern of spots on card C. Information is stored by punching a hole through C at every point where a spot appears. For example, suppose that A has two holes and B three. Six spots will appear on C, some of which may coincide (Figure 6, page 59). On reconstruction of A, 18 rays at most strike S. 6 of these retrace the path of rays which originally passed from A through B onto C, and the other 12 strike S at spurious points which were not originally pinholes in card A (Figure 7, page 60).

Pattern A is consequently seen amongst a background of noise. However, noting that genuine points on S (those which correspond to pinholes on card A) receive three rays while spurious points receive less than three, a threshold detector set at a value of three will produce a faithful reproduction of pattern A.

Other pairs of patterns are then added by punching holes in card C at every point where the new pairs produces a spot. In fact some of these holes will have already been made in the storage of previous pairs.

As the number of pairs of patterns increase, retrieval in the manner outlined above does not remain perfect. For consider the limiting case when the Correlogram stores so much information that it comprises a large number/ . .

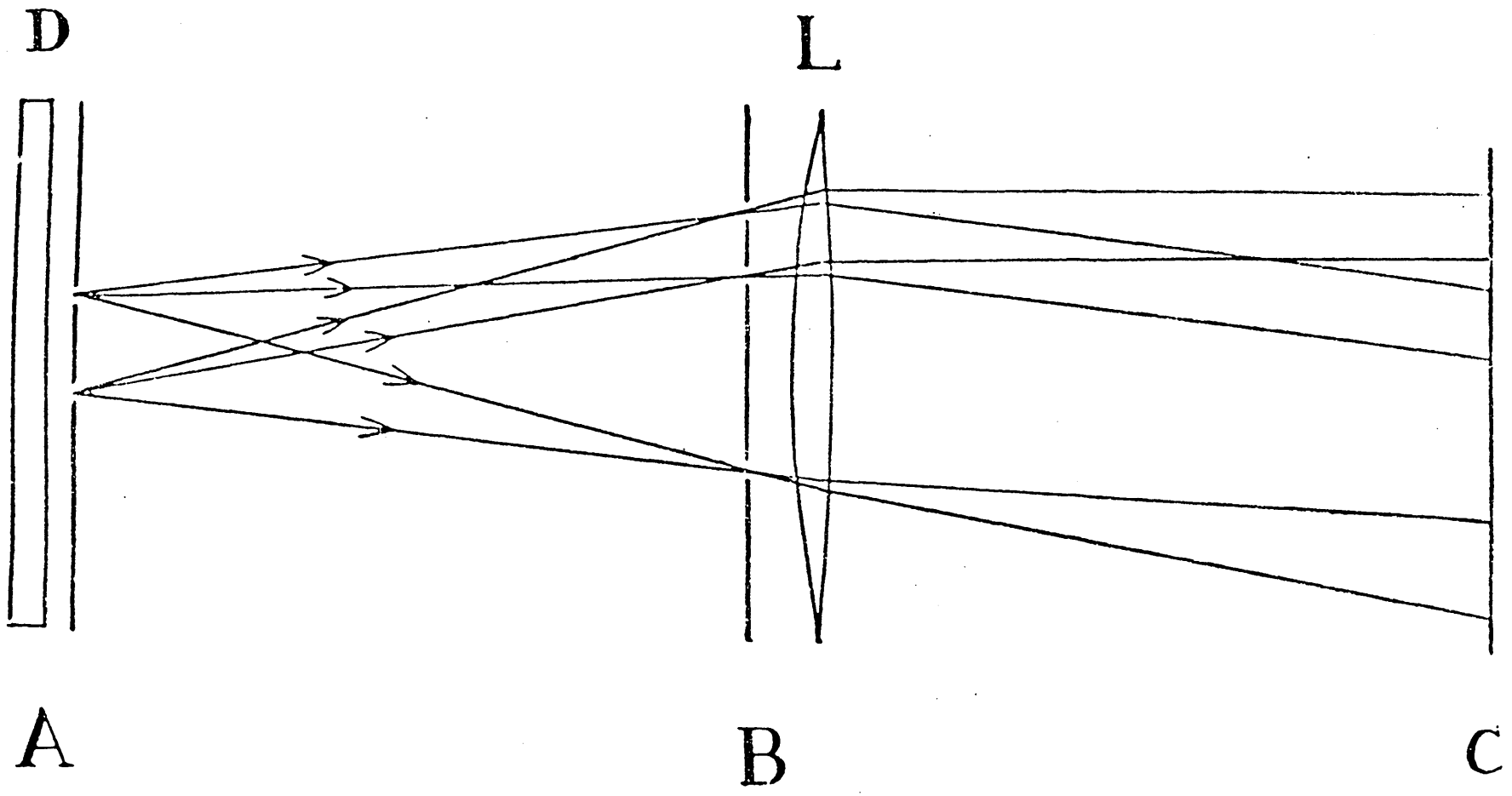

FIGURE 6 - CONSTRUCTING A CORRELOGRAM. D IS A DIFFUSE LIGHT SOURCE, L A LENS AND C THE PLANE OF THE CORRELOGRAM OF A WITH B.

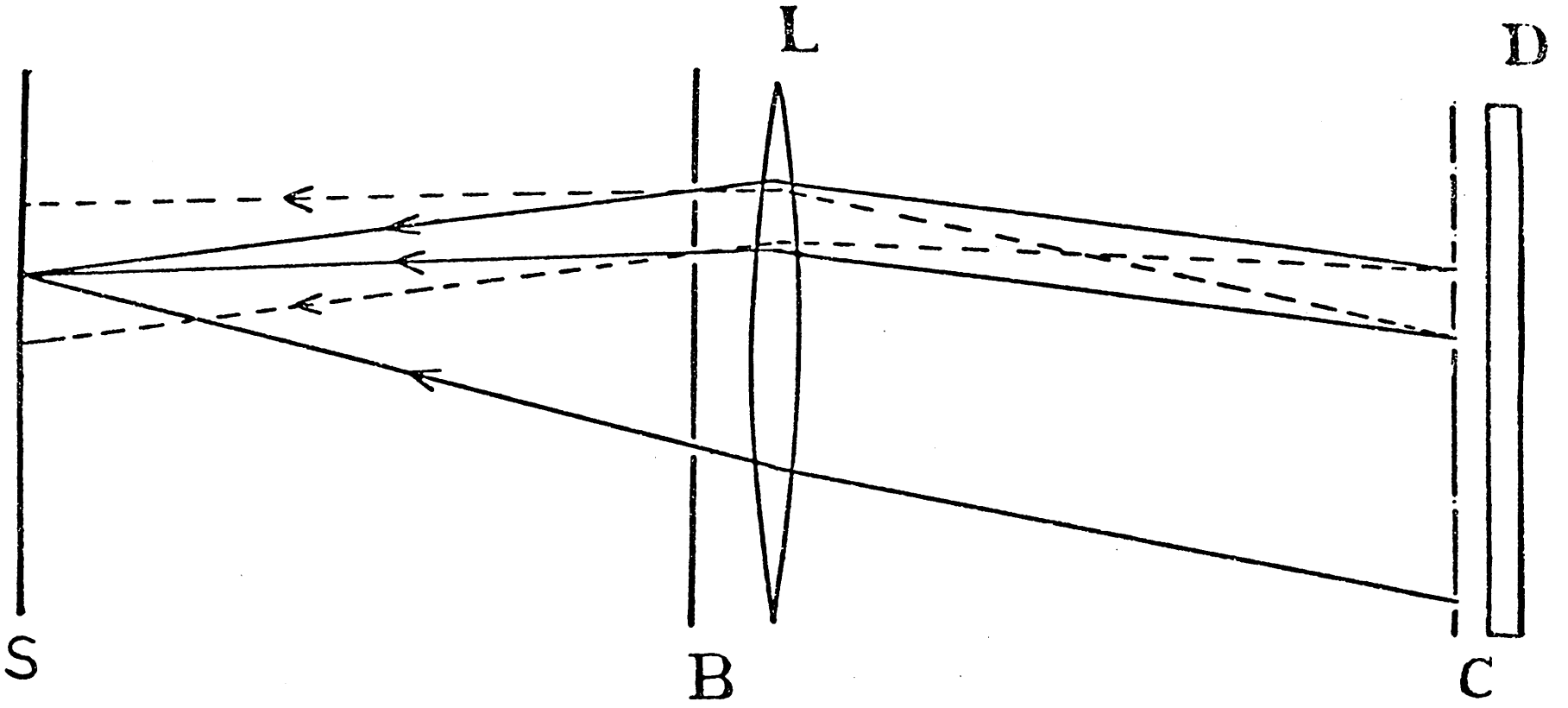

FIGURE 7 RECONSTRUCTING A PATTERN. FULL LINES ARE PATHS TRAVERSED IN FIGURE 6. BROKEN LINES ARE PATHS NOT TRAVERSED IN FIGURE 6. THERE ARE 18 RAYS WHICH TRAVEL FROM C TO S. 5 OF THESE ARE SHOWN HERE.

number of pinholes and nothing else. In a retrieval task, D is uniformly illuminated and the memory performs no useful function.

Analysis

To investigate the conditions for accurate recall the problem will be formulated in discrete terms.

Let each card be divided up into N small sections. The pattern of pinholes on it is representable by an N-dimensional vector. The component A_j, for example, has the value 1 if there is a pinhole at position j on card A, otherwise $A_j=0$. Avoiding such matters as the finite size of the pinholes and diffraction effects and imposing cyclic boundary conditions, a component C_k of C is identified with the N of the N^2 pairs (A_i,B_j) for which $k=j-i$ or $k=j-i+N$. C_k has the value 0 unless at least one of these ordered pairs with which it is connected has the form (1,1), when it is given the value 1. Points on the screen S are similarly identified with points on C and B. A point S_l is identified with the N of the N^2 pairs (C_k,B_j) for which $l=j-k$ or $l=j-k+N$ and is given the value 1 if a certain number of these N pairs, depending on the threshold set, are of the form (1,1). Otherwise $S_l=0$.

We will consider the storage of R pairs of random M-hole patterns, that is M components chosen out of a possible N of a vector representing a pattern to be stored, have the value 1. The other N-M take the value 0.

Consequently / . .

Consequently, a component C_k of card C is identified with RN pairs (A_i, B_j). The probability that it has the value 1 is

$$P_c = 1 - (1 - \frac{M^2}{N^2})^{NR}$$

($\frac{M}{N}$ is the probability that one component of the ordered pair (A_i, B_j) has the value 1)

For $NR \gg 1$, $P_c = 1 - \exp(\frac{-M^2R}{N})$

and so $R = \frac{-N}{M^2} \ln(1-P_c)$ 4.2

P_c is in fact the density of pinholes on the card C.

In the reconstruction process, the threshold must be set as high as possible to catch the genuine spots and to remove as much of the noise as possible. This is achieved if it is set at M. A point S_1 on S which is a spurious point will produce an unwanted spot if it receives M rays through the M holes of B, and this will only happen if the plate C is able to send M rays through B to converge on S_1.

The probability of this occurring is just less than P_c^M.*

The mean number of spurious spots appearing on S is

$$E(\text{spur}) \leq (N-M)P_c^M$$

As more patterns are stored, P_c increases and so does the mean number of errors per pattern retrieved. Certainly, for good recall, this number should be less than 1 and an upper bound on the value of P_c will be placed by setting

---

* Only R-1 stored patterns can contribute to spurious spots on S.

$$E(\text{spur}) = (N-M)P_c^M = 1$$

A slightly safer estimate is

$$NP_c^M = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots 4.3$$

Equations 4.3 and 4.2 enable the information capacity of the system to be deduced. Since the Correlograph retrieves R M-hole patterns with small error then the information gained is

$$I \simeq R\log_2\binom{N}{M} \text{ bits} = R\ln\binom{N}{M} \text{ natural units}^* \text{ (n.u.)}$$

Consider $\ln\binom{N}{M}$. Using Stirling's approximation - namely that for integral n, $\ln(n!) = n.\ln(n) - n$ then

$$\ln\binom{N}{M} = N\ln N - (N-M)\ln(N-M) - M\ln M .$$

Writing $M = \gamma N$,

$$\ln\binom{N}{M} = N\ln N - N(1-\gamma)\ln(N(1-\gamma)) - N\ln\gamma N$$

$$= -N(\gamma \ln\gamma + (1-\gamma)\ln(1-\gamma))$$

For $\gamma \ll 1$, $\ln(1-\gamma) \simeq 0$

then $\ln\binom{N}{M} \simeq -N\gamma\ln\gamma$

$$= M\ln\frac{N}{M} \quad \ldots\ldots\ldots\ldots\ldots 4.4$$

Thus $I = RM\ln\left(\frac{N}{M}\right)$ n.u. for $M \ll N$

$$= N\ln P_c \ln(1-P_c)\left(1 - \frac{\ln M}{\ln N}\right) \quad \ldots\ldots 4.5$$

If, furthermore, lnM is neglected in comparison to lnN then

$$I/N = \ln P_c \ln(1-P_c) \text{ n.u. for } N \longrightarrow \infty$$

Viewed as a function of P_c, this expression has a maximum value at $P_c = \frac{1}{2}$ when

$$I_{MAX}/N = \ln 2 \ln 2 \text{ n.u.}$$

$$=/ \ ..$$

---

* 1 bit = ln(2) natural units of information

$= \ln 2 \triangleq 0.6931$ bits

Fixing P_c at this value determines optimum values of R and M

These are $M = \log_2 N$ and $R = \frac{N}{(\ln N . \log_2 N)}$, and under these conditions the Correlograph is working not so far short of the theoretical maximum for, since C is in fact an N-bit binary register, this memory stores its information about 69% as densely as it can ever do. We define the efficiency E to be the ratio of the total amount of information retrieved from a memory to the theoretical maximum amount of information retrievable from a similar system made up of the same elements. This is a number between 0 and 1 and here can equal 0.69. It should be mentioned that this figure of 69% should be treated as an upper limit to the efficiency which is approached as N increases for the approximations made are only true in the limit of infinite N. The main point is that E is of order unity and can be as much as 0.69.

The behaviour of the efficiency for systems of finite size and when no approximations are made will be discussed in Chapter 5.

The other property of importance that the Correlograph has is that of being able to perform just as well if the pattern it is given is a displaced version of the pattern originally stored. This property is similar to that of the/ . .

the Holophone, whose performance in outputting the whole of a stored signal of which it is only given part is independent of where this part is located in the signal.

Consider all M^2 pairs of points (A_i, B_j) of a particular pair of stored patterns which are of the form (1,1). These define the M^2 rays which impinge on C. One such pair (A_i, B_j) causes the component C_k of C to have the value 1, where k=j-i (modulo N). On reconstruction, if the paths of these rays are retraced, they impinge in groups of M on the M genuine points of the viewing screen. Consider now the vector B' with components $(B_d, B_{d+1}, B_{d+2}, \ldots B_n, B_1, \ldots B_{d-1})$ (cyclic boundary conditions are imposed). We will look at what happens to these same M^2 rays when this displaced pattern is used in recall. For example, the ray passing from C_k through B_{j+d} will produce a spot on S at a position 1 where 1=j+d-k=i+d (modulo N) (k=j-i modulo N) Thus all M^2 genuine rays are shifted sideways by the same amount d. Consequently the pattern A, though displaced, is reproduced from the displaced input B'. Fidelity of reproduction is not altered.

In summary here is a simple distributed model of memory. It functions in parallel, can deal with displaced patterns and can be made to perform efficiently. It was originally conceived, without any biological implementation in mind, to show that it is not necessary to employ such complex systems as holographic memories to/ . .

to perform the tasks they do.

As a candidate for neurophysiological realisation the Correlograph, like holographic models, makes demands on the desired properties of the nervous system. In particular the physical implementation of it mentioned here hides the complex logical operations it performs to be able to deal with displaced inputs. Consider a pair (A_i, B_j) which is identified with an element C_k of C. Suppose the mapping $(A_i, B_j) \rightarrow C_k$ has been effected i.e. C_k has been given the value 1 because some pair of patterns stored had both $A_i=1$ and $B_j=1$. It is easy to see how the inverse mapping $(C_k, B_j) \rightarrow A_i$ is carried out, but the mapping $(C_k, B_{j+d}) \rightarrow A_{i+d}$ presents problems as its inverse mapping $(A_{i+d}, B_{j+d}) \rightarrow C_k$ will not necessarily have been carried out in the storage process.(see Figure 8)

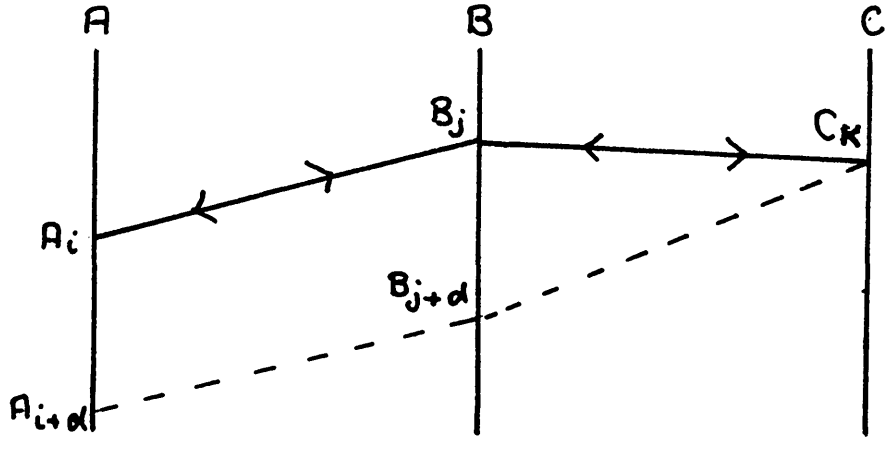

Figure 8

4.4 The Associative Net

A/ . .

Analysis

A model allied to the Correlograph which has a simple logical structure and which, as it cannot deal with displaced patterns, does not suffer from the above disadvantage, is the Associative Net which will now be described.

In this model each ordered pair (A_i, B_j) is identified with a unique component C_k of C. The N_A points of A and the N_B points of B (Now N_A does not equal N_B necessarily) are now represented by a network of N_A parallel lines crossing at right angles N_B parallel lines (Figure 9, page 68). The $N_A N_B$ nodes of the network represent the memory store C. As before, the storage of R pairs of messages will be considered.

A particular pair of messages is stored by sending a pulse down M_A of the N_A A-lines chosen at random and at the same time activating M_B of the N_B B-lines similarly chosen. Node C(i,j) is switched on if it receives a pulse down lines A_i and B_j. Once a node has been switched on it remains on.*

The probability that a particular node has been turned on if R pairs have been stored is

$$P_c = 1 - \left(1 - \frac{M_A M_B}{N_A N_B}\right)^R$$

$$= 1 - \exp\left(-\frac{RM_A M_B}{N_A N_B}\right) \text{ (using the exponential approximation as before)}$$

i.e. R = / . .

* Activated nodes are coloured black in Figure 9.

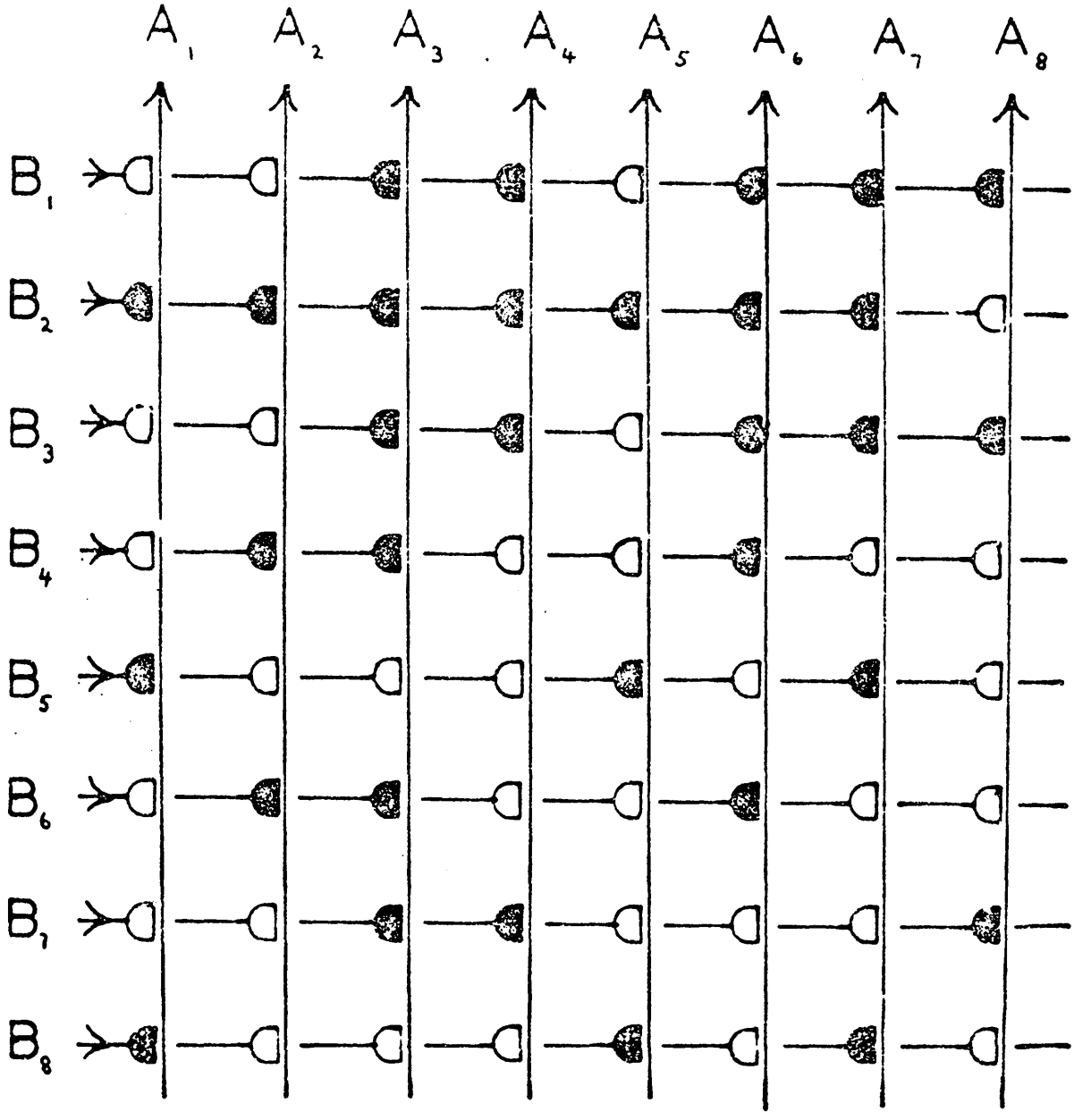

Figure 9 The Associative Net

i.e. $R = \frac{-N_C}{M_A M_B} \ln(1-P_c)$ (where N_C is written for $N_A N_B$)

Using B-messages to recall A-messages, in retrieval the rule is that every node which a pulse in a B-line meets and which has been previously turned on sends a pulse down its A-line. The pulses sent down each A-line feed into a non-linear detector which outputs a pulse if a specified threshold, common to all the output lines, is reached. With the threshold on the A-lines set at M_B then the appropriate A-message will be recalled, accompanied, perhaps, by spurious pulses.

Following the approach used for the Correlograph, the maximum value of P_c is set by the equation

$$N_A P_c^{M_B} = 1 \quad \text{i.e. } M_B = -\frac{\log_2 N_A}{\log_2 P_c} \quad \text{(from 4.3)}$$

The amount of information stored in the memory when R messages have been successfully recalled (disregarding the small error) is

$$I_A = RM_A \log_2 N_A \text{ bits}$$
$$= -\frac{N_C}{M_B} \log_2 N_A \ln(1-P_c)$$
$$= N_C \ln P_c \ln(1-P_c) \text{ n.u.} \quad \ldots\ldots 4.6$$

and $I_{A(max)} = N_C \ln 2$ when $M_B = \log_2 N_A$, $R = \frac{N_A N_B}{M_A M_B} \ln 2$.

The above equations do not specify M_A. However, if it/ . .

it is also required that A-messages should retrieve B-messages with the same small error, then we have the relation $M_A = \log_2 N_B$. Finally, under these conditions the efficiency E is

$$E = \frac{I_{A(max)}}{N_C} = \ln 2 \simeq 0.69,$$

and so the Associative Net can be made to function as efficiently as the Correlograph. We compare these two systems by listing below in Table 1 the optimum parameter settings for an NxN-bit Net and an N-bit Correlograph. We set $N_A=N_B=N$, $M_A=M_B=M$ and regard N as a very large number.

	N-bit Correlograph	NxN-bit Associative Net
P_c	$\frac{1}{2}$	$\frac{1}{2}$
M	$\log_2 N$	$\log_2 N$
R	$\frac{N}{M^2} \ln 2$	$\frac{N^2}{M^2} \ln 2$
E	0.69	0.69

Table 1.

We see that the Associative Net, having more binary registers, stores N times more message pairs than the Correlograph, although it does not have the latter's facility for dealing with displaced inputs. Dispensing with this property is, in fact, an advantage when we consider, as we now do, how the Associative Net may be implemented as / . .

as a neurophysiological system.

A neural representation of the Associative Net

We regard the horizontal lines of figure 9 (page 68) as the axons of the N_B input neurons B_1, B_2, B_3,...B_{N_B}, while the vertical lines are dendrites of the N_A output neurons A_1, A_2, A_3,...A_{N_A}. At each of the $N_A N_B$ axo-dendritic intersections there is a modifiable synapse $C(i,j)$. This becomes facilitated if, in the storage of a particular message pair, neurons A_i and B_j have both been made to fire. On subsequent presentation of a message which causes B_j to fire, the effect of the modified synapse $C(i,j)$ is to depolarise the membrane of the neuron A_i, which then fires if the number of synapses depolarising it reaches a particular threshold value common to all output neurons. Similar synaptic mechanisms have been proposed many times before. This one, involving all-or-none instead of gradual modification of the synapse, is a variation of that proposed by Hebb (1949) which was mentioned in Chapter 2.

Bearing in mind that the human nervous system, for example, contains about 10^{10} nerve cells, it would seem reasonable that a neural Associative Net might have at least 10^6 axons and 10^6 dendrites.

4.5 Non-linearity

We have already seen that the Holophone and the Off-diagonal Holophone are linear models, being described by the equation $\alpha = / \ .\ .$

$$\alpha = M^{-1}DM\beta \qquad (1.1)$$

We shall now show that the Correlograph and the Associative Net are the corresponding non-linear analogues of these models in the sense of the equation

$$\alpha = [[M^{-1}DM]\beta] \qquad (1.2)$$

(This linear/non-linear classification is not quite so clear-cut, as the first two models we looked at deal with messages having components +1 and -1 and the messages of the second two models have components 1 and 0.)

We demonstrate these similarities by considering the storage of R pairs of messages $(A^{(r)},B^{(r)})$ (r=1,2,..R)

The Correlograph

Here we take

$$M_{qs} = \frac{1}{N}\exp(\frac{2\pi i}{N}qs), \quad \phi^A = MA, \quad \phi^B = MB .$$

A and B are NxR matrices, containing information about the R message pairs, with components of value 0 or 1.

As for the Holophone, we choose

$$D_{pq} = \delta_{pq}\sum_{r=1}^{R}\phi^A_{pr}\phi^{B*}_{qr}$$

The non-linear store is thus of the form

$$[M^{-1}DM]_{ms} = \left[\frac{1}{N^4}\sum_r\sum_{p,q,l,j}\delta_{pq}\, e^{\frac{2\pi i}{N}(-mp+pl-qj+qs)}A_{lr}B_{jr}\right]$$

$$= \left[\frac{1}{N^3}\sum_r\sum_{l,j}\delta_{l,m-s+j}\,A_{lr}B_{jr}\right]$$

= / . .

$$= \left[\frac{1}{N^3} \sum_r \sum_j A_{m-s+j,r} B_{jr} \right]$$

(limits of summation are omitted in many cases)

This is an NxN matrix containing only N independent components, for all those which have the same value of m-s are identical. Each component with suffix j-(m-s+j) = s-m has a value depending on the RN pairs $(A_{m-s+j,r}, B_{jr})$ (r=1,2,...R and j=1,2,...N) (All arithmetic operations are performed modulo N), and thus can be identified with the component C_{s-m} of the Correlogram (page 61) We can thus write equation 1.2 as

$$\alpha_m = \left[\sum_m C_{s-m} \beta_s \right].$$

Changing the notation to be consistent with that used in the earlier discussion of the Correlograph, (i.e. we make the substitutions l=m and j=s), for an input $B^{(n)}$ the output α_ℓ depends on the N pairs $(C_k, B_{j1}), (C_k, B_{j2}), \ldots$ (C_k, B_{jN}) where k=j-l. This can be identified with the output from the Correlograph.

Thus, by a particular choice of the matrix D, equation 1.1 describes the <u>linear</u> Holophone, equation 1.2 the <u>non-linear</u> Correlograph.

<u>The Associative Net</u>

A is now an N_AxR matrix and B an N_BxR matrix. The Fourier Transform matrices are N_AxN_A or N_BxN_B matrices as required. As for the Off-diagonal Holophone, D is defined / . .

defined as

$$D_{pq} = \sum_{r=1}^{R} \phi^{A}_{pr} \phi^{B*}_{qr}$$

We now have

$$[M^{-1}DM]_{ms} = \left[\frac{1}{N_C^2} \sum_{r} \sum_{p,q,l,j} e^{\frac{2\pi i}{N_A}(pl-mp)} e^{\frac{2\pi i}{N_B}(qs-qj)} A_{lr}B_{jr} \right]$$

$$(N_C = N_A N_B)$$

$$= \left[\frac{1}{N_C} \sum_{r} \sum_{l,j} \delta_{lm}\delta_{js} A_{lr}B_{jr} \right]$$

$$= \left[\frac{1}{N_C} \sum_{r} A_{mr}B_{sr} \right] \quad \ldots\ldots\ldots 4.7$$

The value of this component of the non-linear store depends on the R pairs (A_{mr}, B_{sr}) $(r=1,2,\ldots.R)$ and represents the state of node C(m,s) of the Associative Net. For the retrieval process we now write $B^{(n)}$ for β in equation 1.2. Then

$$\alpha_m = \left[\sum_{s} [M^{-1}DM]_{ms} B_{sn} \right] \ldots\ldots 4.8$$

and the value of the output α_m depends on the summation of activity in all nodes of output line A_m of the Associative Net.

The relationship between the linear Off-diagonal Holophone and the non-linear Associative Net is thereby established for, with D chosen as above, equation 1.1 describes the former and equation 1.2 the latter.

Conclusion

It / . .

Conclusion

It has been shown in this chapter that the tasks which two particular holographic models of memory can perform can be carried out by two much simpler models. Secondly, it has been possible to make analytical statements about the performance of non-linear representations of both models in the limit when they are infinitely large.

It is proposed in the next chapter to present an account of the properties of these models, taking particular notice of their finite size. Where analytical results have not been obtained use is made of the results of computer simulation experiments.

CHAPTER 5

The Correlograph and the Associative Net. 2

5.1 Introduction

The properties of the Associative Net and the Correlograph will now be discussed in some detail. We shall see how they perform when faced with damage to their store locations or when inaccurate addresses are employed (It will be recalled that we have indicated that these tasks are ones with which local stores find difficulty in dealing). We shall also look at the models acting as content-addressable memories, when the address is itself an arbitrary part of the information to be retrieved (as in the Holophone), and finally consider how efficiently they may perform when the values of the message components are drawn from a binomial ensemble.

Since the mathematics of the two devices are very similar, emphasis will be placed on the Associative Net, with reference where necessary to the Correlograph.

5.2 Associative Nets of finite size

Firstly, however, it is necessary to say a little more about the approximations made in Chapter 4 in obtaining the conditions under which the Net performs best.

The results derived are only strictly true in the limit of infinite N_A and N_B. This is because the messages are not retrieved accurately, each containing

on / . .

on average one extra pulse, and the information required to identify these unwanted pulses has been neglected.

Consider one A-line of the Associative Net. With B-messages being used to recall A-messages, either one pulse or no pulse is output in each retrieval task. If a pulse is output then this might be a spurious or a genuine pulse. If no pulse is output then no such confusion arises.

If, as before, R pairs of messages are stored in the Net then, with the usual nomenclature, we have

$$R = \frac{-N_A N_B}{M_A M_B} \ln(1-P_c) \quad \ldots\ldots\ldots 5.1$$

and the probability of a spurious line firing is

$$P_S = P_c^{M_B} \quad * \quad \ldots\ldots\ldots\ldots\ldots 5.2$$

The probability that an A-line should output a 1 (that is a pulse) is

$$P_A = \frac{M_A}{N_A} .$$

It should output a 0 (no pulse) with the probability $1-P_A$. What does in fact happen is that it outputs a 1 (a spurious 1 or a genuine 1) with probability

$$\delta = P_A + (1-P_A)P_S$$

The probability of outputting a 0 is $1-\delta = (1-P_A)(1-P_S)$

---

* Unless R is very small it is safe to make this equality.

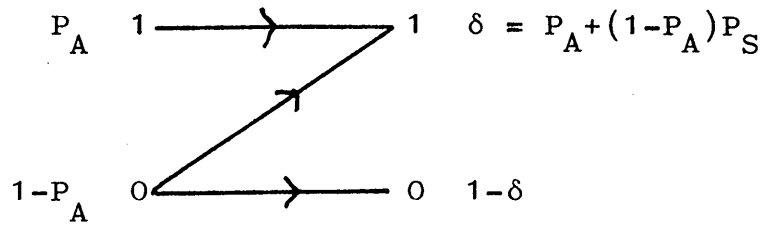

Figure 10

The state transition diagram is shown in Figure 10. Each A-line can be viewed as a communication channel which distorts what <u>should</u> be output (with probabilities shown on the left) to produce what <u>is</u> output (with probabilities shown on the right).

The values of the four conditional probabilities will be examined. P(a/b) represents the probability that a <u>should</u> have been output when b <u>was</u> output. Since if a pulse is not output, no distortion has taken place,

$$P(0/0) = 1 \quad \text{and} \quad P(1/0) = 0$$

Furthermore P(1/1) is equal to the probability that a 1 <u>should</u> be output and <u>is</u> output, divided by the probability that a 1 <u>is</u> output.

$$\text{So} \quad P(1/1) = \frac{P_A}{\delta} \quad \text{and also} \quad P(0/1) = 1 - \frac{P_A}{\delta}$$

If retrieval is flawless then the information gained per retrieval task by observing the output of the A-line being considered is

$$I_O = -(P_A \ln P_A + (1-P_A)\ln(1-P_A)) \text{ n.u. . . .} \quad 5.3$$

However, since the message is not reconstructed perfectly, information is still required to identify the inaccuracies/ . .

inaccuracies present in it and has the value

$$\Delta I_0 = -P(1)(P(1/1)\ln P(1/1) + P(0/1)\ln P(0/1))$$
$$-P(0)(P(1/0)\ln P(1/0) + P(0/0)\ln P(0/0)) \text{ n.u.}$$
$$= -\delta\left(\frac{P_A}{\delta}\ln\frac{P_A}{\delta} + \left(1-\frac{P_A}{\delta}\right)\ln\left(1-\frac{P_A}{\delta}\right)\right) \text{ n.u.}$$

with $\delta = P_A + (1-P_A)P_S$.

Writing $fP_A = P_S$ and $g = (1-P_A)f$,

then $\delta = (1+g)P_A$, $\quad \frac{P_A}{\delta} = \frac{1}{1+g}$,

$$1-\frac{P_A}{\delta} = \frac{g}{1+g}, \qquad \delta - P_A = gP_A$$

Consequently

$$\Delta I_0 = -P_A \ln\left(\frac{g}{1+g}\right) - gP_A \ln\left(\frac{g}{1+g}\right)$$
$$= P_A(1+g)\ln(1+g) - gP_A \ln f - gP_A\ln(1-P_A)$$
$$\ldots\ldots\ldots 5.4$$

From 5.3 and 5.4, the precise expression for the amount of information gained is

$$I = I_0 - \Delta I_0$$
$$= -P_A\ln P_A - (1-P_A)\ln(1-P_A) + gP_A\ln(1-P_A)$$
$$+ gP_A\ln f - P_A(1+g)\ln(1+g)$$
$$= -P_A\ln P_A + gP_A\ln f - (1-P_Af)(1-P_A)\ln(1-P_A)$$
$$- P_A(1+g)\ln(1+g) \ldots\ldots\ldots 5.5$$

This is the information gained from one line in one retrieval / . .

retrieval task. The information gained from the N_A lines when used to retrieve the R stored messages is

$$I_T = RN_AI \quad \ldots\ldots\ldots\ldots 5.6$$

Now $R = -\frac{N_AN_B}{M_AM_B}\ln(1-P_c)$,

where $M_A = P_AN_A$ and $M_B = \frac{\ln P_S}{\ln P_c} = \frac{\ln fP_A}{\ln P_c}$ (from 5.1 and 5.2)

Hence, from 5.5 and 5.6, the efficiency E is given by

$$E = \frac{I_T}{N_AN_B}$$

$$= \frac{\ln P_c \ln(1-P_c)}{P_A \ln fP_A}\Big(P_A \ln P_A - gP_A \ln f + (1-P_Af)(1-P_A)\ln(1-P_A) + P_A(1+g)\ln(1+g)\Big) \text{ n.u.}$$

Maximising E with respect to its arguments P_A, P_c and f will not only determine the conditions under which the memory will perform optimally but may also show whether it is more efficient in information theoretical terms to retrieve a lot of messages inaccurately or fewer messages more accurately.

The quantity E is the product of the functions

$$F(P_c) = \ln P_c \ln(1-P_c) \quad \text{and}$$

$$G(P_A,f) = \frac{1}{P_A \ln fP_A}\Big(P_A\ln P_A + (1-P_Af)(1-P_A)\ln(1-P_A) - gP_A\ln f + P_A(1+g)\ln(1+g)\Big). \quad \ldots\ldots 5.7$$

where $g = (1-P_A)f$.

The behaviour of E was investigated by treating $G(P_A,f)$ / . .

$G(P_A,f)$ and $F(P_c)$ independently. On that assumption, if E is to be as large as possible then P_c must certainly be chosen so that $F(P_c)$ takes its maximum value.

This is ensured by setting $P_c=0.5$, when $F(P_c)=(\ln 2)^2$. (Although P_c, f and P_A are related by the equation $P_c^{M_B}=fP_A$, it will be seen that choosing P_c in this way does not restrict the choice of f and P_A and hence the value of $G(P_A,f)$ as there is some freedom in the choice of M_B.)

The behaviour of $G(P_A,f)$ was not amenable to study by analytical methods and a program was written in ALGOL to tabulate the values of this function for a range of values of f at various fixed values of P_A. These results are plotted out in Figure 11 on page 80.

Since $F(P_c)$ is, in fact, the function used to calculate the information efficiency by the previous approximate methods of Chapter 4, the value of $G(P_A,f)$ compared with 1 is a measure of the validity of those approximations used.

As f is determined by the equation

$$P_S = fP_A = P_c^{M_B}$$

and M_B may have any value between 1 and N_B, then, with P_A fixed, f is restricted to have a positive value up to P_c/P_A when $M_B = 1$, and so the range of possible values of G is constrained by the value of P_c. However, unless P_A is very large, (that is equal to or

greater / . .

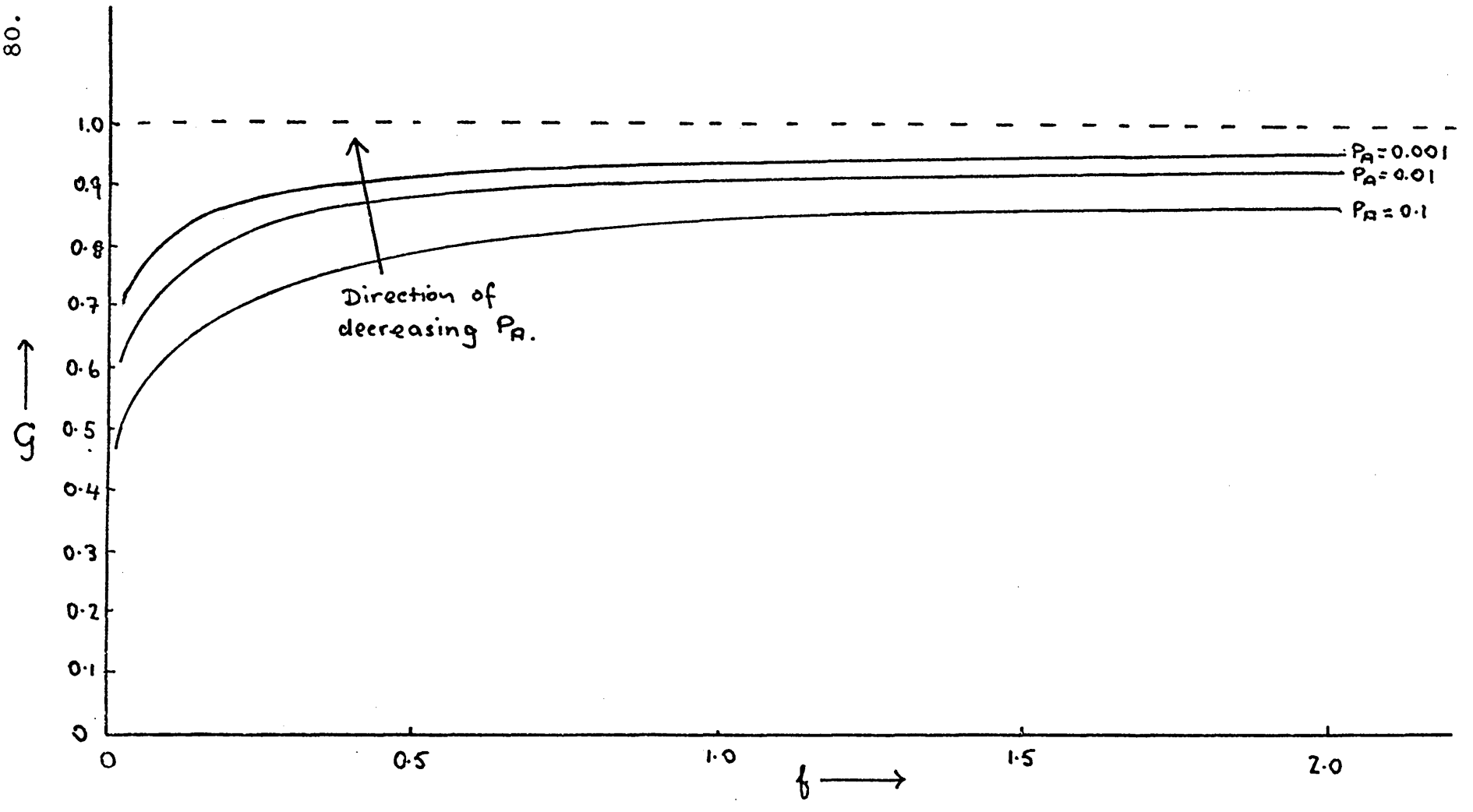

FIGURE 11. $G(P_A, f)$ plotted as a function of f, P_A being fixed

greater than P_c) the upper limit $G(P_A, P_c/P_A)$ to $G(P_A,f)$ occurs on the almost horizontal section of the curve of G plotted as a function of f and is a very slowly changing function of P_c. Thus $G(P_A,f)$ and $F(P_c)$ can be treated independently.

When $f=P_c/P_A$ and $P_c=\frac{1}{2}$, from 5.7 (substituting also $g=(1-P_A)f$, then the upper limit to G at constant P_A is

$$G(P_A,f) = -\frac{1}{P_A \ln 2}\left(P_A \ln P_A + \tfrac{1}{2}(1-P_A)\ln(1-P_A) - \tfrac{1}{2}(1-P_A)\ln\frac{1}{2P_A} + P_A\left(1+\frac{1-P_A}{2P_A}\right)\ln\left(1+\frac{1-P_A}{2P_A}\right)\right)$$

$$= 1 - \frac{1}{2P_A \ln 2}\left((1-P_A)\ln(1-P_A)+(1+P_A)\ln(1+P_A)\right)$$

This increases monotonically from 0 to 1 as a function of P_A. For P_A very small

$$G\left(P_A, \frac{1}{2P_A}\right) = 1 - \frac{1}{2P_A \ln 2}\left(-(1-P_A)P_A + (1+P_A)P_A\right)$$

$$= 1 - \frac{P_A}{\ln 2}$$

which tends to 1 as P_A tends to 0.

Recalling that the efficiency of retrieving information is $E = F(P_c).G(P_A,f)$, then it seems that, for a given value of P_A and P_c, E increases as the probability of error fP_A does so and the value of f for which E is the greatest is P_c/P_A when the Net is working with address messages each comprising a single pulse ($M_B=1$)

However, this may be explained by noting that the expression for the number of pairs stored R is

$$R = /\ .$$

$$R = - \frac{\ln P_c \ln(1-P_c) N_A}{P_A \ln fP_A}$$

At constant P_A, as f increases R also does so. Furthermore, since the probability of a spurious pulse occurring in an output line is $P_S = fP_A$, then as f increases, the information retrieved per message decreases.

The conclusion is that the memory functions best in information theoretical terms if a large number of messages is stored but with only a small amount of information being able to be extracted on retrieval of any one from store.

As far as realisable Associative Nets are concerned, the value of f, and hence of M_B, chosen will depend on the required fidelity of retrieved information. The smaller the value allowed of the ratio of the number of bits of information gained in retrieving a message to the number of bits contained in that message, then the better the performance of the Net. Since the value of $G(P_A,f)$ can never exceed 1, it can never be used with an information efficiency greater than ln2= 0.69.

Values of $G(P_A,f)$ are tabulated below in Table 2 when the memory is functioning under the conditions found to be optimum in the approximate derivations of Chapter 4.

Here $N_A P_c^{M_B} = 1$. Since $P_S = fP_A = P_c^{M_B}, N_A P_A f = 1$.

i.e. / . .

$$\text{i.e. } f = \frac{1}{P_A N_A} = \frac{1}{M_A} .$$

It must be stressed that for a given size of Net the efficiency E depends on the value of f chosen. For example - choosing f to equal 1 instead of $\frac{1}{M_A}$ increases the maximum value of E. Thus the values of $G(P_A,1)$ are also tabulated for comparison.

N_A	$M_A=\frac{\log_2 N_A}{\log_2 P_c}$	$P_A=\frac{M_A}{N_A}$	$G(P_A,\frac{1}{M_A})$	$G(P_A,1)$	$1-\frac{\ln M_A}{\ln N_A}$
2^4	4	2.5×10^{-1}	0.63	0.76	0.5
2^8	8	3.1×10^{-2}	0.73	0.89	0.63
2^{10}	10	9.8×10^{-3}	0.76	0.92	0.67
2^{15}	15	4.6×10^{-4}	0.81	0.95	0.76
2^{20}	20	1.9×10^{-5}	0.84	0.97	0.78

Table 2 $P_c=0.5$

Conclusion

It has been shown that, by judicious choice of P_c and M_B, the information efficiency of the Net may have any value up to ln(2). In particular, if the values of these parameters are chosen so that $P_c=0.5$ and the mean number of errors per message retrieved is 1, then, as long as the Net is large enough ($N_A = N_B > 2^{10}$), we see from Table 2 that the maximum information efficiency obtainable is within 76% of the limiting value of ln(2). Since under these conditions the efficiency is relatively high, we shall employ this mathematically useful criterion of allowing / . .

allowing 1 error per message in subsequent analysis. We note that the previous approximate treatment, which arrived at a maximum efficiency of ln(2) by using this criterion, gave a too optimistic result. A more realistic estimate is obtained by approximating the quantity $\ln\binom{N_A}{M_A}$ to $M_A \ln(\frac{N_A}{M_A})$ (equation 4.4), instead of neglecting $\ln M_A$ compared to $\ln N_A$. This leads to an expression for the efficiency of

$$E = \ln P_c \ln(1-P_c)(1 - \frac{\ln M_A}{\ln N_A}) \quad \text{n.u.}$$

Values of the function $1 - \frac{\ln M_A}{\ln N_A}$ with $M_A = \log_2 N_A$ are also shown in Table 2.

Attention will now be turned to the problem of the behaviour of the Associative Net in the face of damage. We distinguish the two cases of damage to store locations and distortion of the addresses (or cues) used.

5.3 Damage to the store elements

Suppose that the Net, which has $P_c N_A N_B$ nodes turned on, has a fraction 1-q of these turned off at random. The number of pulses that a genuine line emits in recall is not now M_B but a number drawn from a binomial distribution of mean qM_B. Consequently, the threshold on each output line must be lowered to a value tM_B (where t is less than 1). Not only, however, must a genuine line fire with large probability, but also there must / . .

must only be a small probability that a spurious line, down which on average $P_c qM_B$ pulses travel, must fire. These two conditions are set out below, and the two relevant probability distribution curves are shown below in Figure 12.

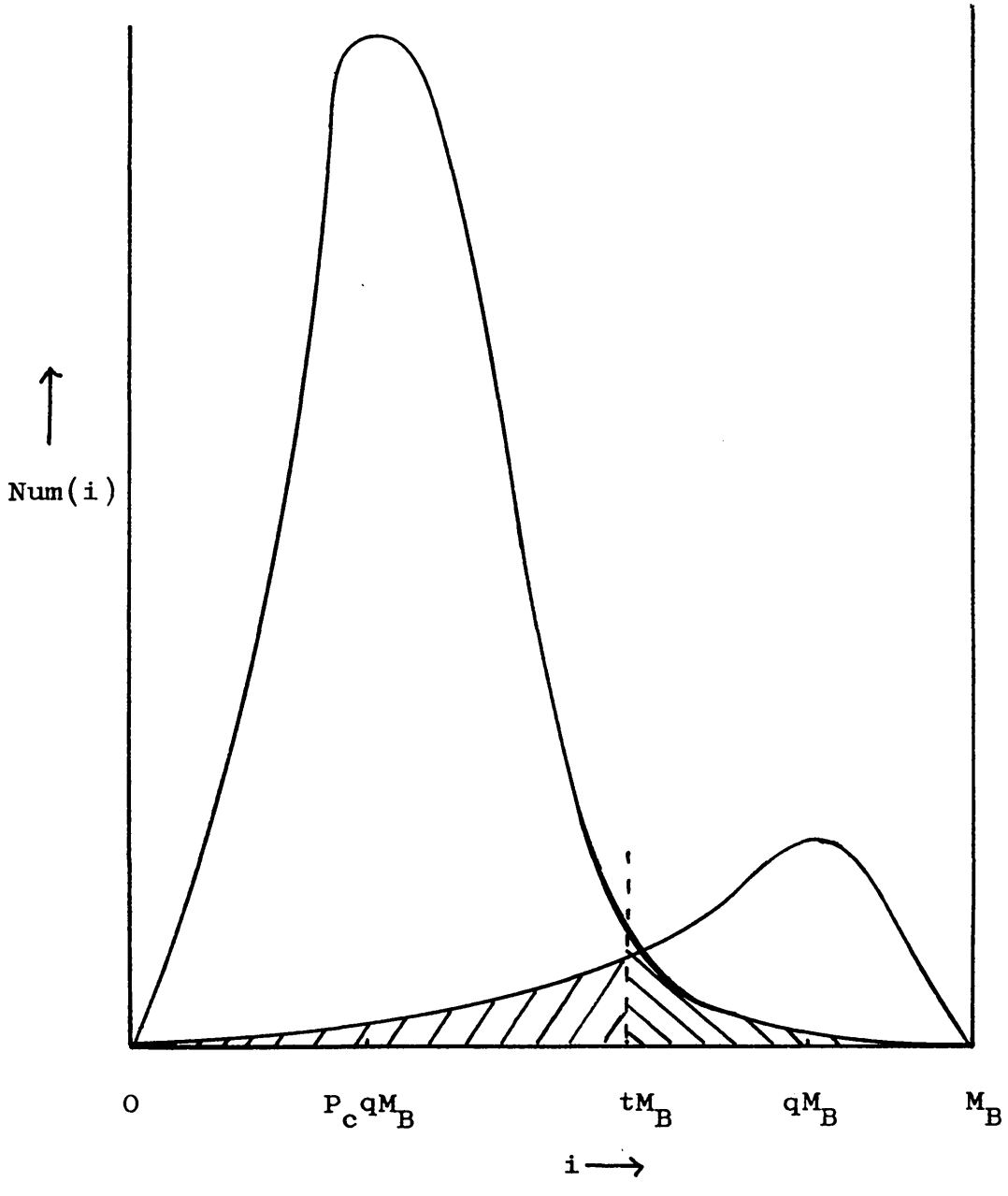

Figure 12 (not to scale)

Here / . .

Here the steeply rising curve shows the number Num(i) of spurious output lines down which i pulses travel in the recall of one message from store; the other curve relates to the genuine lines. Consequently, with a threshold at tM, the shaded area to the right of the vertical line drawn at $i=tM_B$ is proportional to the number of spurious lines which finally output a pulse and the shaded area to the left is proportional to the number of genuine lines which do <u>not</u> finally output a pulse. The ratio of the area under the spurious curve to that under the genuine curve is $(N_A-M_A)/M_A$.

With the threshold at t, the probability that a genuine line will not fire is

$$\sum_{i=0}^{M_B t-1} \binom{M_B}{i} q^i (1-q)^{M_B-i}$$

The probability that a spurious line will fire is

$$\sum_{i=tM_B}^{M_B} \binom{M_B}{i} (P_c q)^i (1-P_c q)^{M_B-i}$$

Using the criterion that for good recall of a message the number of missing genuine pulses and the number of extra spurious pulses must not each exceed 1, then we have

$$M_A \cdot \sum_{i=1}^{M_B t-1} \binom{M_B}{i} q^i (1-q)^{M_B-i} \leqslant 1 \quad \ldots\ldots\ldots \quad 5.8$$

and

$(N_A-M_A)/$ . .

$$(N_A - M_A) \sum_{i=M_B t}^{M_B} (P_c q)^i (1-P_c q)^{M_B - i} \leqslant 1 \quad \ldots\ldots 5.9$$

5.8 and 5.9 are certainly true if

$$M_A \sum_{i=1}^{M_B t} \binom{M_B}{i} q^i (1-q)^{M_B - i} \leqslant 1 \quad \ldots\ldots 5.10$$

and

$$N_A \sum_{i=M_B t}^{M_B} \binom{M_B}{i} (P_c q)^i (1-P_c q)^{M_B - i} \leqslant 1 \quad \ldots\ldots 5.11$$

Since both are sums of tails of binomial distributions, we can replace to a good approximation the logarithm on the left hand sides of 5.10 and 5.11 by the logarithm of the largest term of each.

i.e. $\ln M_A + \ln \left(\binom{M_B}{M_B t} q^{M_B t} (1-q)^{M_B (1-t)} \right) \leqslant 0$

and $\ln N_A + \ln \left(\binom{M_B}{M_B t} (P_c q)^{M_B t} (1-P_c q)^{M_B (1-t)} \right) \leqslant 0$

The upper limit to the loading of the Net is obtained by regarding these relationships as equalities. Using Stirling's approximation in writing

$$\ln \binom{M_B}{M_B t} = -M_B \left(t \ln t + (1-t) \ln (1-t) \right), \text{ then}$$

$$\frac{\ln M_A}{M_B} = t . \ln \frac{t}{q} + (1-t) \ln \frac{1-t}{1-q} \quad \ldots\ldots 5.12$$

and

$$\frac{\ln M_A}{M_B} \quad / \ldots$$

$$\frac{\ln N_A}{M_B} = t \ln \frac{t}{P_c q} + (1-t)\ln \frac{1-t}{1-P_c q} \quad \ldots\ldots 5.13$$

Now the damaged Associative Net retrieves information with an efficiency

$$E \simeq \frac{R \ln \binom{N_A}{M_A}}{D_{THEORY} N_A N_B}$$

where D_{THEORY} is the maximum amount of information (in natural units) which can ever be gained from a binary register damaged to this extent.

With $R = -\frac{N_A N_B}{M_A M_B} \ln(1-P_c)$ (equation 5.1)

and writing $\ln \binom{N_A}{M_A} \simeq M_A \ln \left(\frac{N_A}{M_A} \right)$ (equation 4.4) then

$$E \simeq \frac{-\ln(1-P_c) \ln \binom{N_A}{M_A}}{D_{THEORY} M_B}$$

By subtracting 5.12 from 5.13,

$$\frac{1}{M_B} \ln \left(\frac{N_A}{M_A}\right) = t \ln \frac{1}{P_c} + (1-t)\ln \frac{1-q}{1-P_c q} \quad , \text{and so}$$

$$E \simeq \ln(1-P_c) \left(t \ln P_c + (1-t) \ln \left(\frac{1-P_c q}{1-P_c} \right) \right) / D_{THEORY} \quad \ldots 5.14$$

(As we saw at the end of section 5.2 the fact that the message is not perfectly retrieved can be offset by the approximation to $\ln \binom{N_A}{M_A}$ used.).

We can obtain an exact expression for D_{THEORY}. This is the maximum amount of information retrievable from a binary/ . .

binary switch for which, if it has been turned on, there is a probability 1-q that it has been turned off accidentally. Let the switch have been turned on with probability p. When it is subsequently examined, the information gained from it is

$$D_o = -(p\ln p + (1-p)\ln(1-p)) \text{ natural units of information} \quad . \; . \; . \quad 5.15$$

There is a probability 1-q, however, that it has been switched off. The state transition diagram is shown below in Figure 13.

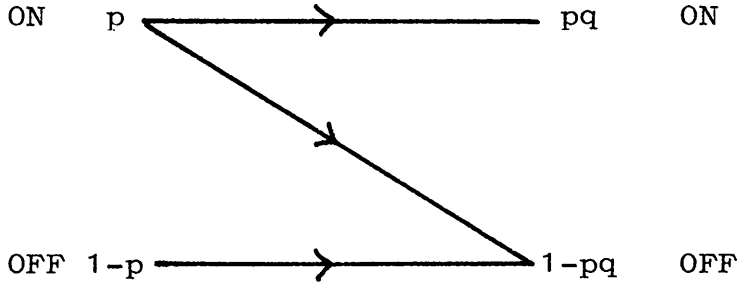

Figure 13

Ambiguity only arises if the switch is off. The information required to clear up this uncertainty is

$$\Delta D = -(1-pq)\left(\frac{1-p}{1-pq}\cdot\ln\frac{1-p}{1-pq} + \frac{p(1-q)}{1-pq}\cdot\ln\frac{p(1-q)}{1-pq}\right)$$

$$= -\Big((1-p)\ln(1-p) + p(1-q)\ln(p(1-q)) - (1-pq)\ln(1-pq)\Big) \text{ n.u.} \quad . \; . \; . \quad 5.16$$

Thus, from 5.15 and 5.16, the information gained from this switch is

$$D = D_o - \Delta D = -pq\ln p + p(1-q)\ln(1-q) - (1-pq)\ln(1-pq) \text{ n.u.} \quad . \; . \; . \quad 5.17$$

Partially/ . .

Partially differentiating D with respect to p,

$$\frac{\delta D}{\delta p} = -q\ln p - q + (1-q)\ln(1-q) + q + q\ln(1-pq)$$

$$\frac{\delta D}{\delta p} = 0 \text{ when } q\ln(1-pq) + (1-q)\ln(1-q) = q\ln p \quad . . \quad 5.18$$

$$\text{i.e. } (1-pq)^q(1-q)^{(1-q)} = p^q,$$

$$\frac{p}{1-pq} = (1-q)^{\frac{1-q}{q}},$$

$$\frac{1}{p} = q + (1-q)^{-\frac{1-q}{q}} \quad . . . . . \quad 5.19$$

The second partial differential of D is

$$\frac{\delta^2 D}{\delta p^2} = -\frac{q}{p} - \frac{q^2}{1-pq}$$

$$= -q\left(\frac{1}{p} + \frac{q}{1-pq}\right) < 0$$

since p and q are numbers having values between 0 and 1. Consequently equation 5.18 gives the maximum value of p. Calling this value p_m, then from 5.17

$$D_{THEORY} = D(p_m) = p_m q \ln p_m + p_m(1-q)\ln(1-q) - (1-p_m q)\ln p_m + (1-p_m q)\left(\frac{1-q}{q}\right)\ln(1-q)$$

$$\text{i.e. } D_{THEORY} = \ln\left(\frac{1}{p_m}\right) + \frac{1-q}{q}\ln(1-q) \quad \text{natural units} \quad . . 5.20$$

$$\text{where } \frac{1}{p_m} = q + (1-q)^{-\frac{(1-q)}{q}}$$

Thus D_{THEORY} can be calculated directly, given a value of q.

The equations describing the performance of the damaged/ . .

damaged Net were investigated by choosing $N_A = N_B = N = 10^6$, setting $M_A = M_B = M$ and proceeding as follows:

A particular value of q was selected and equation 5.12 was used to obtain a functional relationship between t and M. This enabled a functional relationship between t and P_c to be found by means of 5.13, q and N being fixed. Finally, equation 5.14 was used to find the value of the efficiency E as a function of P_c, with D_{THEORY} calculated from 5.20. The maximum values of E for a range of values of q are listed in Table 3 together with appropriate values of q, t, P_c, M, R and D_{THEORY}.

q	t	P_c	M	R	D_{THEORY}	E_{MAX}
0.4	0.261	0.224	111	2.1×10^7	0.246	0.12
0.5	0.340	0.240	84	3.9×10^7	0.322	0.14
0.75	0.557	0.257	43	1.6×10^8	0.588	0.18
0.9	0.731	0.3	31	3.7×10^8	0.763	0.24
0.95	0.811	0.337	27	5.6×10^8	0.857	0.27
1.00	1.0	0.45	17	2.0×10^9	1.0	0.55

Table 3 $N_A=N_B=N=10^6$, $M_A=M_B$

(The fact that the value of E_{MAX} for q=1 is less than $\ln 2 \simeq 0.69$ is due to the finite size of the net.)

Conclusion / . .

Conclusion

It can be seen from Table 3 that the Associative Net sacrifices efficiency of information storage for the ability to work reliably if damaged. This is accomplished by loading the system more lightly ($P_c \simeq \frac{1}{4}$) than in the undamaged case, that is by storing fewer pairs of messages, each of which must have relatively large values of M. We have considered damage to be a destructive process only as, for the sake of neurophysiological plausibility, we have allowed only active nodes (synapses) to be damaged by becoming inactive. This is also the reason for choosing N to have the value 10^6 in the arithmetic calculations.

5.4 The effect of using inaccurate addresses

Retrieval of information by means of an inaccurate cue (or address) is not logically equivalent to using a damaged store. In the latter case a different number of pulses may be sent down each input line, whereas in the former situation it is known that at least that number of pulses equal to the number of active input lines which should be active are transmitted down each genuine output line. This remark corrects the previous assertion made concerning the logical equivalence of retrieval from a damaged store and retrieval using incomplete cues (Willshaw and Longuet-Higgins, 1970)

We consider the retrieval of a particular A-message from / . .

from a square Associative Net, where once again we set $M_A = M_B = M$. We suppose that each B-message contain T ones (i.e. activates T lines), sT of which should be present. Furthermore, this number is only a fraction g of the total number M of ones which the cue should have. We thus have $gM = sT \quad 0 \ll g \ll 1, \quad 0 \ll s \ll 1$. It is assumed that the threshold on the A-lines can be set at a value sT, thus ensuring that all genuine A-lines fire. The probability that a spurious A-line will fire is

$$\sum_{i=0}^{i=(1-s)T} \binom{T}{sT+i} P_c^{sT+i} (1-P_c)^{(T-sT)-i} ,$$

and, as before, we determine the limiting value of P_c by the equation

$$N\binom{T}{sT} P_c^{sT} (1-P_c)^{(T-sT)} = 1, \text{ (provided } P_cT \ll sT\text{)}.$$

Thus, by using Stirling's approximation

$$\ln N + T\left(s \ln \frac{P_c}{s} + (1-s)\ln\left(\frac{1-P_c}{1-s}\right)\right) = 0 \quad . \quad . \quad .5.21$$

Since $M = \frac{sT}{g}$,

$$M = \frac{-s \ln N}{g\left(s \ln \frac{P_c}{s} + (1-s)\ln\left(\frac{1-P_c}{1-s}\right)\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad 5.22$$

In this case we shall define the efficiency E of the system to be the ratio of the information retrieved from it to that which can ever be retrieved from N^2 binary registers / . .

registers. We do this because of the difficulty of defining, for an arbitrary memory system, what constitutes an address whose damage is specified by the parameters g and s.

$$\text{Thus } E = \frac{R\log_2\binom{N}{M}}{N^2}$$

$$\simeq \frac{-\ln(1-P_c)\log_2\left(\frac{N}{M}\right)}{M} \quad \ldots\ldots\ldots \quad 5.23$$

(from 4.4 and 5.1)

Giving N the value 10^6 (as before), for a particular pair of values of g and s, equation 5.22 was used to find a functional relationship between M and P_c. The maximum value $E_{MAX}(g,s)$ of E viewed as a function of P_c was then calculated from 5.23. Values of $E_{MAX}(g,s)$ are shown in Table 4, together with the appropriate values of gM, M, T, R and P_c. gM is the number of genuine ones present in the cue, M is the number that should be present and T is the total number of ones present. In the table these three quantities are rounded off to the nearest integer.

Table 4/ . .

g	s	gM	M	$T=\frac{M}{s}$	R	P_c	E_{MAX}
1	1	17	17	17	2.0×10^9	0.45	55%
0.9	0.9	16	17	19	1.2×10^9	0.3	32%
0.8	0.8	16	20	25	7.0×10^8	0.24	22%
0.7	0.7	15	21	30	4.0×10^8	0.18	14%
0.6	0.6	16	26	43	2.0×10^8	0.15	9%
0.5	0.5	16	32	64	1.2×10^8	0.12	6%
0.9	0.5	16	18	36	4.0×10^8	0.12	11%
0.5	0.9	16	31	35	3.6×10^8	0.3	17%
0.8	0.4	16	20	50	2.4×10^8	0.09	7%
0.4	0.8	16	40	50	1.7×10^8	0.24	10%

Table 4 $N_A=N_B=N=10^6$, $M_A=M_B=M$.

Conclusion

If it is known that the Net is to receive cues incorrect to the extent specified by a pair of values of g and s then the messages stored must have their value of M such that the quantity gM has a particular value only dependent on the size N of the Net. Then, by loading the system lightly it can be made to function relatively efficiently. The lower part of Table 4 shows that the presence of extra pulses in the cue has

a/ . .

a more severe effect than the absence of pulses that should be present. We note that one of the conditions which must hold and does (Table 4) is that P_c must be substantially less than s. This is required in the derivation of equation 5.21 by approximating the tail of a binomial distribution.

5.5 The Auto-Associative Net

Introduction

Little has been said about the performance of the Associative Net when it is required to retrieve information given an incomplete cue which is in fact part of the message to be recalled. This is the task that the Holophone is required to perform. The analogous task - when the Associative Net stores pairs of messages of the form (A,A) instead of (A,B) - will now be considered.

We will investigate the behaviour of a square Net for which $N_A = N_B = N$ and $M_A = M_B = M$. In the storage of a particular pair of messages, the same pattern of pulses is sent along the A-lines and along the B-lines. Let an incomplete previously stored message containing L (L<M) pulses be input along the B-lines. The task is to retrieve the rest of it by looking at the output from the A-lines.

With the threshold at L, as before, the limiting value/ . .

value of P_c is obtained by setting

$$NP_c^L = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 5.24$$

Now L of the M pulses of the message are known, so that M-L remain to be found. Consequently the information gained on successful retrieval of a message is

$$\ln\binom{N-L}{M-L} \text{ n.u.}$$

The total amount of information gained in retrieving all R messages in that way is

$$I = R \ln\binom{N-L}{M-L} \text{ n.u.}$$

$$- \frac{-N^2}{M^2} \ln(1-P_c)(M-L)\ln\left(\frac{N}{M-L}\right) \text{ n.u.}$$

(from 5.1 $R=\frac{-N^2}{M^2}\ln(1-P_c)$ and from 4.4 we write $\ln\binom{N}{M-L} \approx (M-L)\ln\frac{N}{M-L}$)

From 5.24, $\ln N = -L.\ln P_c$. Thus, writing $L = \beta' M$

$$\frac{I}{N^2} = \beta'(1-\beta')\ln P_c \ln(1-P_c) \frac{\ln\frac{N}{M-L}}{\ln N} \text{ n.u.}$$

By maximising $\frac{I}{N^2}$ with respect to β' and P_c, which are independent variables, in the limit as N approaches infinity when $\frac{\ln\frac{N}{M-L}}{\ln N}$ approaches 1, the memory works most efficiently when

$$P_c = 0.5 \quad \text{and} \quad \beta' = 0.5.$$

Hence $\frac{I}{N^2} = \frac{1}{4}(\ln 2)$ bits, and since $L = \log_2 N$ (from 5.24)

M = / . .

$$M = \frac{L}{\beta'} = 2\log_2 N$$

and

$$R = \frac{1}{4}\left(\frac{N}{\log_2 N}\right)^2 \ln 2.$$

Since the pattern of nodes that have been turned on is symmetrical about the leading diagonal of the net, for large N the number of independent binary registers employed is

$$\frac{N^2+N}{2} \simeq \frac{N^2}{2}$$

Consequently the information efficiency for large N is

$$E \simeq \frac{\frac{1}{4} \ln 2 \; N^2}{\frac{N^2}{2}} = \tfrac{1}{2} \ln 2 \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; 5.25$$

This figure is smaller by a factor of 2 than in the case when 'cross-correlations' are being stored (section 4.4). This is because in an "Auto-Associative Net" any one of $\binom{M}{L}$ different messages are allowed to serve as cues whereas the "Cross-Associative Net" associates just one input with one output. If it were known that the L ones of the cue were always related to a particular area of store then only a fraction $\frac{L(M-L)}{M^2}$ $= \beta'(1-\beta')$ of the store need be considered (Figure 14, page 99).

The information efficiency in this case is

E = / . .

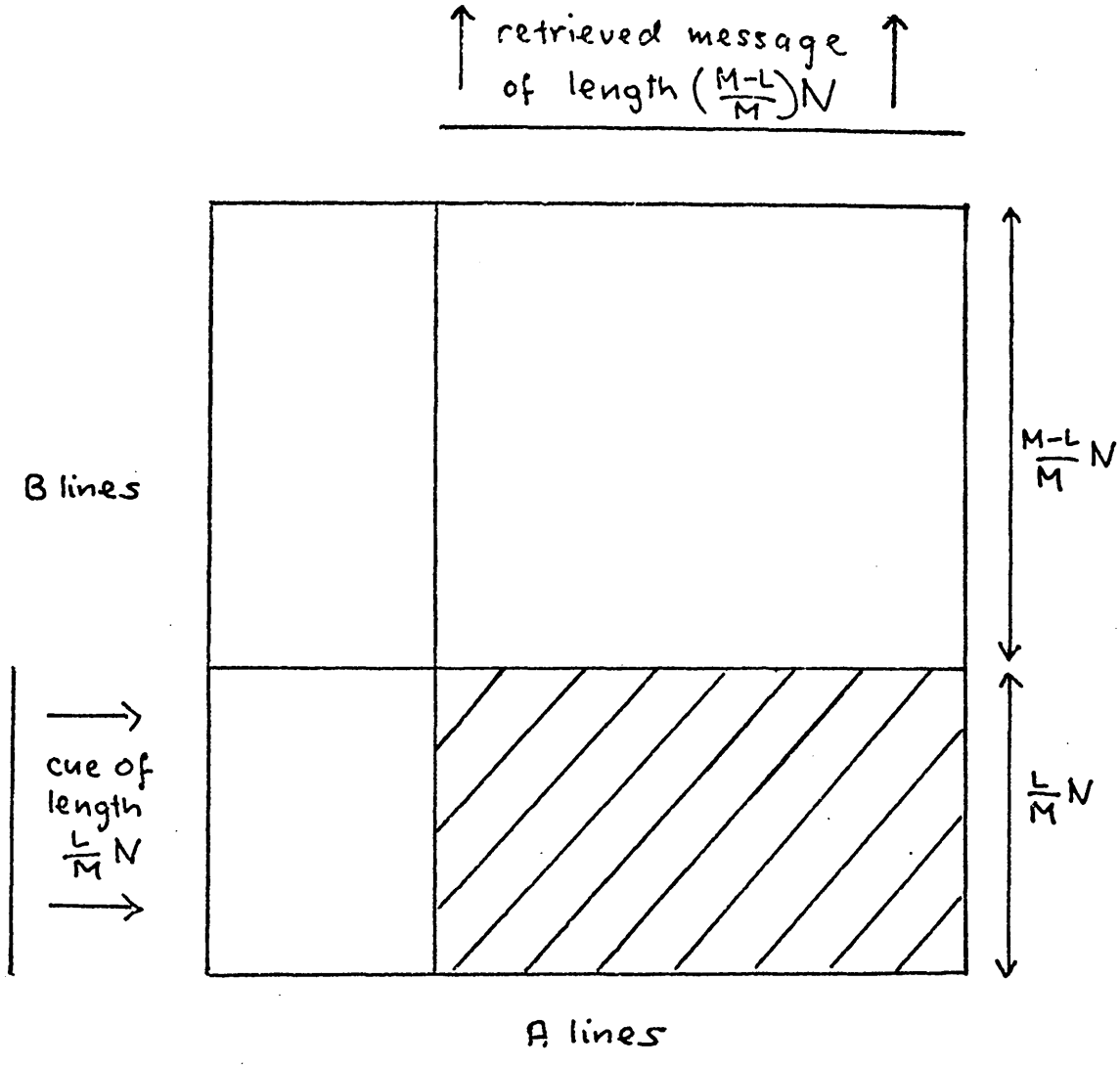

FIGURE 14

An Auto-Associative Net.

If the cue is always a particular section of a stored message then only the shaded area of the store is ever used.

$$E \triangleq \frac{\beta'(1-\beta')\ln P_c \ln(1-P_c)\ln\frac{N}{M-L}}{\beta'(1-\beta')\ln N} \quad \text{n.u.}$$

and has a value for N large of

$\ln P_c \ln(1-P_c)$ n.u., which has a maximum of ln2 bits when $P_c = 0.5$ - independent of β'. The values of M and R are $M = \frac{\log_2 N}{\beta'}$, $R = (\frac{N}{\beta' \log_2 N})^2 \ln 2$. In fact, here the Net is functioning as a Cross-associative Net.

We conclude that the Auto-associative Net sacrifices information efficiency for freedom in the choice of possible cues.

5.6 Storage of messages selected from a binomial distribution

So far the analysis of the properties of the Associative Net has assumed that the values of the parameters M_A and M_B, which specify the number of pulses in the A-messages and in the B-messages respectively, are the same for each of the R message pairs stored. There are an infinite number of ways in which these parameters could be assumed to vary and what will be considered now is the situation when, on storing a message, a pulse is sent down an A-line with probability P_A and a pulse is sent down a B-line with probability P_B. Summing over the R pairs of messages stored, the mean number of ones in an A-message is $\mu_A = P_A N_A$. Similarly for the B-message we define $\mu_B = P_B N_B$.

By / . .

By analogy to 5.1 we can write

$$R = -\ln(1-P_c)\frac{N_A N_B}{\mathcal{N}_A \mathcal{N}_B} \quad \ldots\ldots\ldots\ldots 5.26$$

$$= \frac{-\ln(1-P_c)}{P_A P_B} \qquad \text{provided } P_A \ll 1 \text{ and } P_B \ll 1.$$

We look at a particular A-line for which, in the retrieval of one particular message, no pulse should be output. The probability that a spurious pulse is received is now not the same whatever A-message is retrieved but depends on the number of pulses in the B-message which is acting as the cue. Let this activate m_B input lines. The probability that a spurious A-line is activated is $P_c^{m_B}$.

Summing over all R pairs of stored messages, since the probability that a B-message associated with a given A-message activates m_B input lines is

$$\binom{N_B}{m_B} P_B^{m_B}(1-P_B)^{N_B-m_B}$$

then the mean number of spurious pulses per retrieved A-message which contains m_A genuine pulses is

$$\sum_{m_B=1}^{N_B} (N_A-m_A)\, P_c^{m_B} \binom{N_B}{m_B} P_B^{m_B} (1-P_B)^{N_B-m_B}$$

As before, it will be assumed that accurate retrieval is obtained if the mean number of spurious pulses for one retrieved message, summed over all messages, / . .

messages, is not greater than 1.

Thus $$\sum_{m_B=1}^{N_B} (N_A-m_A)(P_cP_B)^{m_B}(1-P_B)^{N_B-m_B} \leqslant 1$$

Proceeding as before by dropping the term in m_A compared to that in N_A and converting this expression into an equality to place an upper bound on the value of P_c, then we have

$$N_A(1-P_B(1-P_c))^{N_B} = 1$$

$$\text{i.e. } \exp(-P_BN_B(1-P_c)) = \frac{1}{N_A} \quad \ldots\ldots\ldots\ldots 5.27$$

Consider the information gained by retrieving a particular A-message which contains m_A ones. There are $R\binom{N_A}{m_A}P_A^{m_A}(1-P_A)^{N_A-m_A}$ A-messages which contain m_A ones present in store.

The information gained is

$$\ln\binom{N_A}{m_A} \text{ n.u.}$$

and the mean amount of information gained in retrieval of the R stored messages is

$$I = R\sum_{m_A=1}^{N_A}\binom{N_A}{m_A}P_A^{m_A}(1-P_A)^{N_A-m_A}\ln\binom{N_A}{m_A} \text{ n.u.}$$

$$\simeq R\sum_{m_A=1}^{N_A} m_AP_A^{m_A}(1-P_A)^{N_A-m_A}\ln\left(\frac{N_A}{m_A}\right) \text{ n.u. (using 4.4)}$$

Assuming that the variation in $\ln m_A$ can be neglected, if we write $\ln m_A = \ln \mu_A$ this series can be summed / . .

summed.

Thus $$I \simeq RP_A N_A \ln\left(\frac{N_A}{\mu_A}\right) \text{ n.u.} \quad \ldots\ldots\ldots 5.28$$

Now from 5.27 and 5.28,

$$R = \frac{-\ln(1-P_c)}{P_A P_B}$$

and $$P_B = \frac{\ln N_A}{N_B(1-P_c)} \quad ,$$

so $$R = \frac{-(1-P_c)\ln(1-P_c)N_B}{P_A \ln N_B}$$

Substituting in 5.28, the efficiency E is finally given as

$$E = \frac{I}{N_A N_B} \simeq -(1-P_c)\ln(1-P_c)\,\frac{\ln\frac{N_A}{\mu_A}}{\ln(N_A)} \text{ n.u. per bit}$$

For large N_A, E has a maximum at $P_c = 1 - \frac{1}{e} \simeq 0.63$

i.e. $E_{MAX} = \frac{1}{e}$ n.u. $= \frac{\log_2 e}{e} \simeq 0.53$ bits per bit

By reference to 5.27, P_B is set at

$$P_B = \frac{\ln N_A . e}{N_B} \quad \text{and writing} \quad \gamma = \frac{e}{\log_2 e} \simeq 1.89,$$

$$\mu_B = P_B N_B = \gamma \log_2 N_A .$$

P_A remains unspecified. If, however, the Net were required to work in both directions (A-messages retrieving B-messages as well) then P_A would be such that

$$\mu_A = P_A N_A = \gamma \log_2 N_B$$

It follows from 5.26 that in this case

R = / . .

$$R = \frac{N_A N_B}{\gamma^2 \log_2 N_A . \log_2 N_B}$$

Conclusion

In this situation, where the parameters M_A and M_B, which are the number of pulses in the A-messages and in the B-messages, have values drawn from independent binomial distributions, the Net works most efficiently ($E_{MAX} \simeq 0.53$) when the mean number of pulses μ_A and μ_B are approximately twice the respective values M_A and M_B have in the case when they are fixed Further, since here the Net is loaded just over half-full, the number of stored messages R is reduced by a factor of 4. This calculation of the maximum efficiency is not exact, since the parameters of the Net were set so that on average each retrieved message contains one error and (as has been done before) the information required to identify this error has been neglected. Since some A-messages are retrieved more reliably than others (because all the B-messages which act as cues do not activate the same number of input lines), the exact information theoretical calculation is not simple and will not be attempted All that can be said is that the conditions just derived apply in the limit as the size of the Net becomes infinitely large, when the information required to clear up the remaining uncertainty / .

uncertainty becomes neglible.

5.7 Application of the results of Chapter 5 to the Correlograph

The conclusions reached in sections 5.2 - 5.6 about the performance of the Associative Net under certain conditions apply almost without exception to the Correlograph.

The main difference relates to the analysis of section 5.5. For the Auto-correlograph the effective number of independent binary registers making up its store cannot be reduced if it is known from what part of the message to be retrieved the cue originates. This is because each of the Auto-correlogram's N binary registers can receive information from every component of the messages to be stored and thus the Auto-correlograph can cope with a larger variety of cues (displaced inputs produce displaced outputs). Consequently, the maximum information efficiency of the Auto-correlogram is $\frac{1}{2}\ln 2$ instead of $\ln 2$. Otherwise the results obtained for an NxN Associative Net apply to an N-bit Correlograph, except that for the Correlograph the number of pairs of stored messages R is consistently a factor of N smaller than that for the Associative Net.

5.8 Computer simulation experiments

It / . .

It was thought to be valuable to test some of the ideas put forward in the last two chapters by computer simulation. Since similar mathematics underlie their behaviour there was no need to test both models in this way, and so the Correlograph was simulated using the computer language POP-2. It was chosen in preference to the Associative Net in order that the value of N used be as large as possible, for the approximations used in some of the theory are only valid in the limit of infinite N.

In all simulation studies N was set at 1024 and the values of the components of the stored messages were chosen by means of a pseudo-random number generator.

The theory derived in Chapter 4 shows that if N=1024 the system works best with $P_c = 0.5$, $M = \frac{-\log N}{\log P_c} = 10$, $R = \frac{-N}{M^2} \ln(1-P_c) = 7$; that is, when 7 pairs of messages, each message containing 10 out of 1024 components of value 1 are stored, with a single component of the Correlogram equally likely to have the value 1 or 0. This is provided that one error per message retrieved is allowed.

(i) The first experiment was designed to show that as more message pairs with M=10 are stored, then the value of P_c which is such that the mean number of errors per message retrieved is one, defines the limiting performance of / . .

of the system. For P_c greater than this value the number of errors per message increases sharply, setting P_c just less than this value ensures that very few errors in recall occur. In Figure 15 (see page 109) is plotted the mean number of errors made in recall as a function of P_c, the mean being computed over all messages in store. A-messages are used to recall B-messages and vice versa.

(ii) Of course, if the limiting value of P_c were chosen to be other than 0.5, as long as M were also changed so that the relation $NP_c^M = 1$ still held then, by plotting the mean number of errors per message retrieved as a function of P_c, a similar sharply increasing curve would be obtained. The second experiment was designed to show that, of the limiting values of P_c that could be chosen, if the value of 0.5 were selected the system performs most efficiently.

Pairs of M and R were chosen by selecting a value of P_c (say P_c'), computing $m=-\log N/\log P_c'$ and $r=-N\ln(1-P_c')/m^2$ and since these numbers are not necessarily integral, choosing M and R to be the nearest integers to m and r.

R pairs of messages with M=10 were then stored and retrieved (The density of spots P_c in the correlogram is not necessarily equal to P_c'.) This was repeated for values of P_c in the range 0 to 1 and the information gained in retrieval plotted as a function of P_c in Figure 16/ . .

Figure 16 (see page 110). Two curves are shown. The upper one refers to the information stored, the lower one to the information gained in retrieval. It can be seen that both curves have a maximum at a value of P_c very close to but not equal to 0.5. This discrepancy and the fact that the maximum information efficiency obtainable is not 69% is because N is not infinitely large.

(iii) It has been argued in section 5.2 that, as the system is loaded up by storing more pairs of messages, so that more than one error per message is made in recall, although the information retrieved per message decreases, the total number of bits gained in retrieval can increase as a function of P_c up to a value not greater than Nln2 bits. With M fixed, pairs of messages were added one by one to the store and after each addition the information gained for the retrieval of all messages was calculated. Here, two simulations were performed, with M taking in turn the value 5 and then 10. It will be seen from Figures 17A and 17B (see pages 111 and 112) that an efficiency of ln2 = 69% is almost reached but never exceeded.

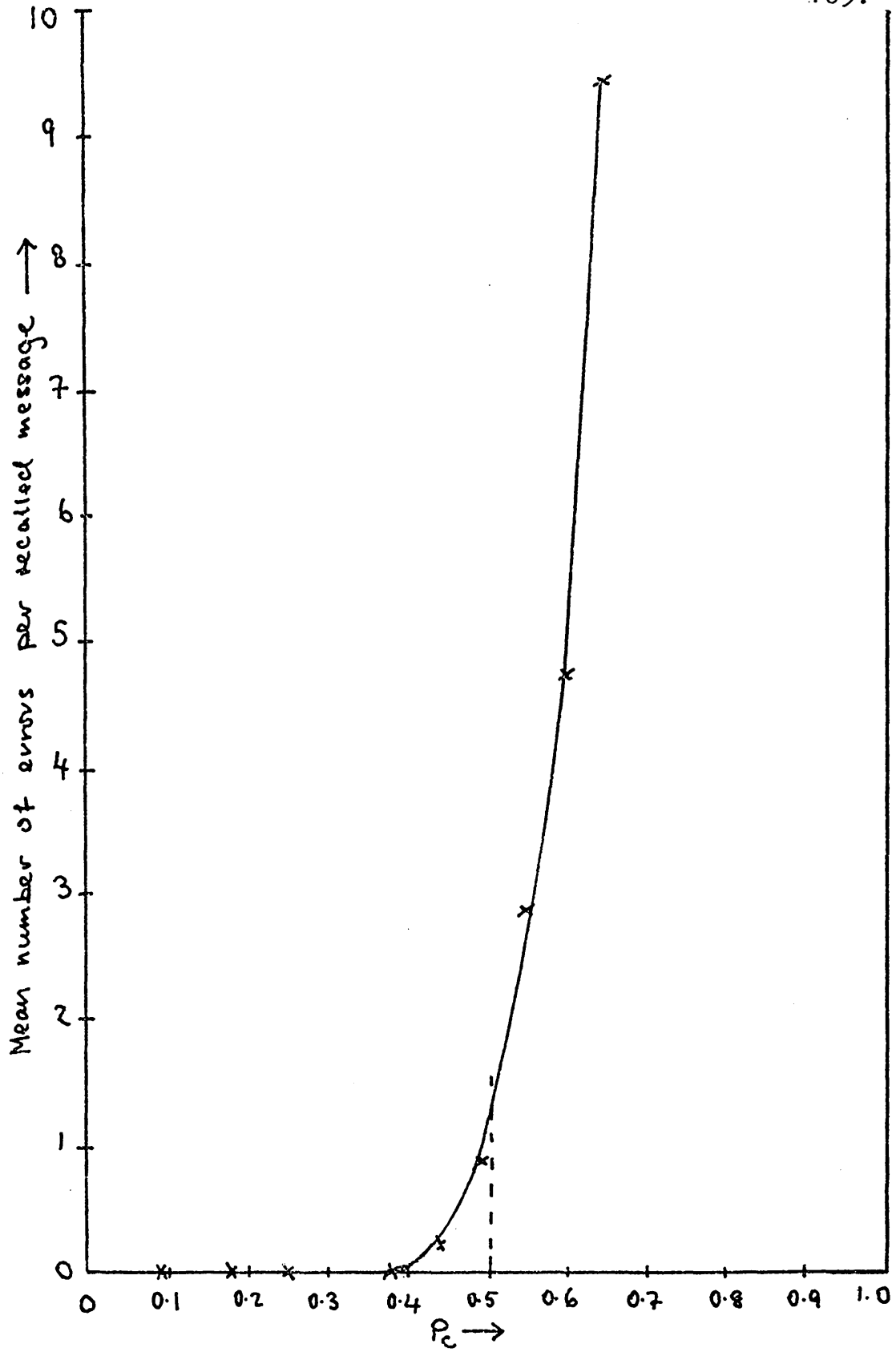

FIGURE 15 N = 1024, M = 10.

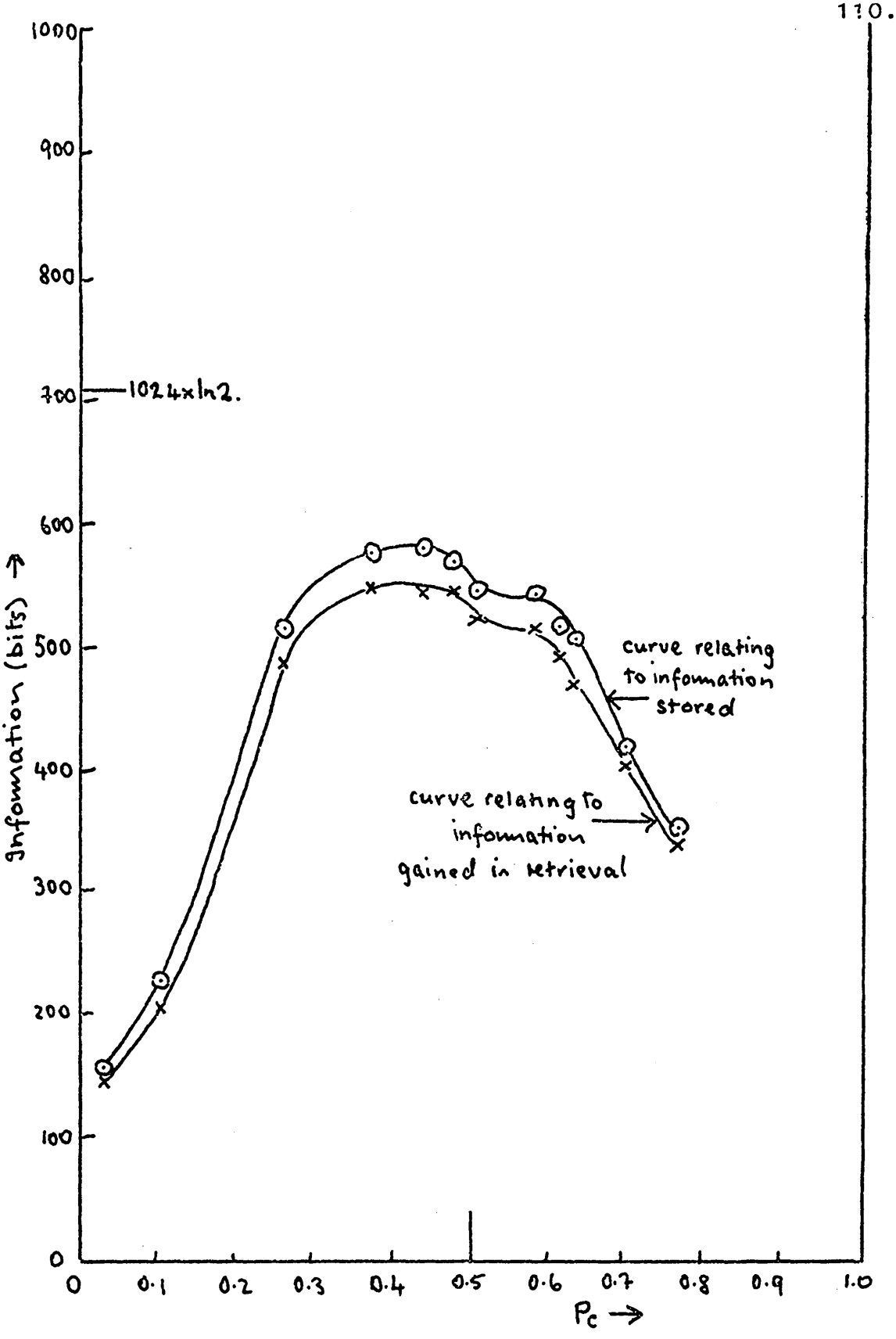

FIGURE 16 $N = 1024$. M and P_c are such that $NP_c^M = 1$.

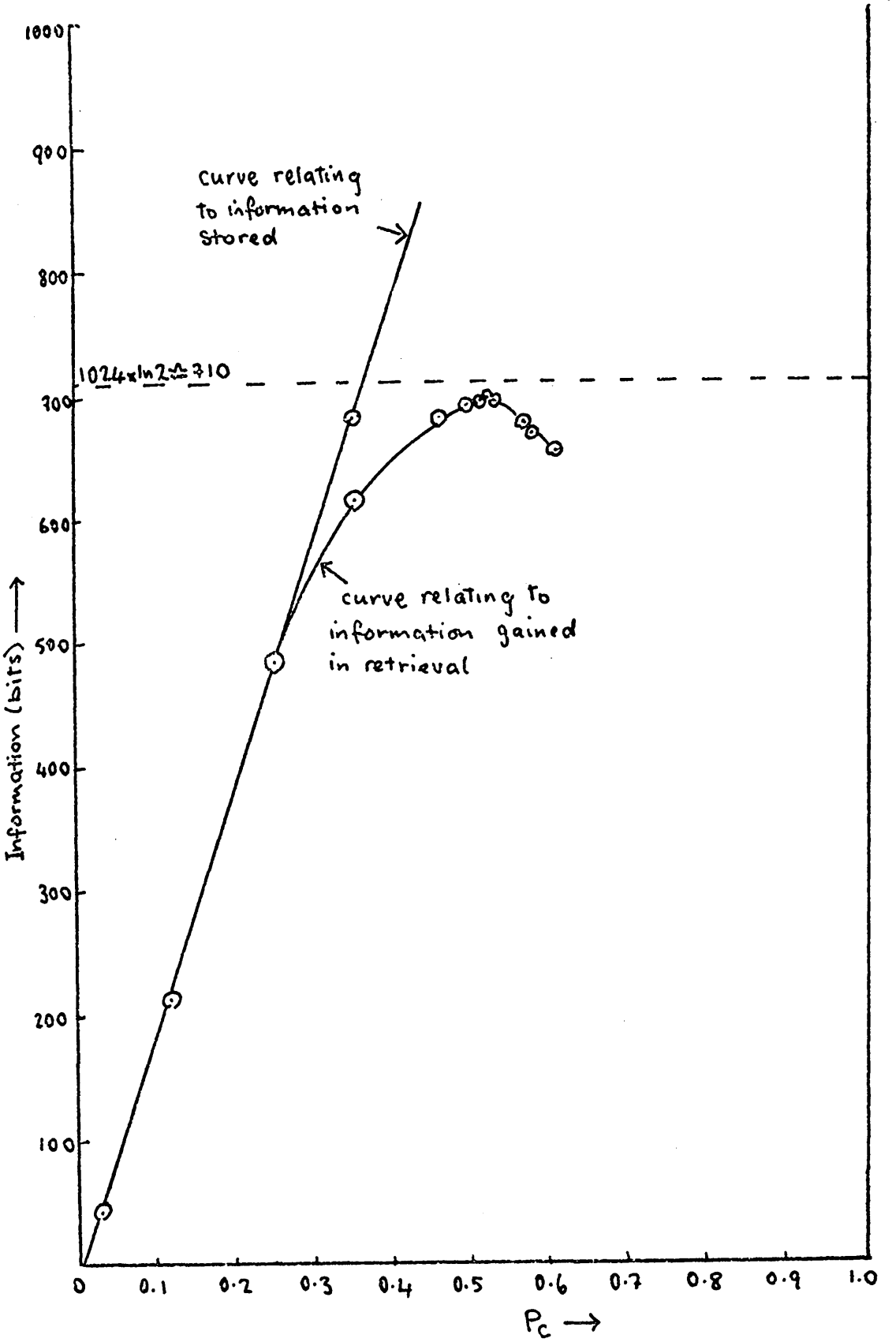

FIGURE 17A N=1024, M=5.

$NP_c^M = 1$ at $P_c = 0.25$.

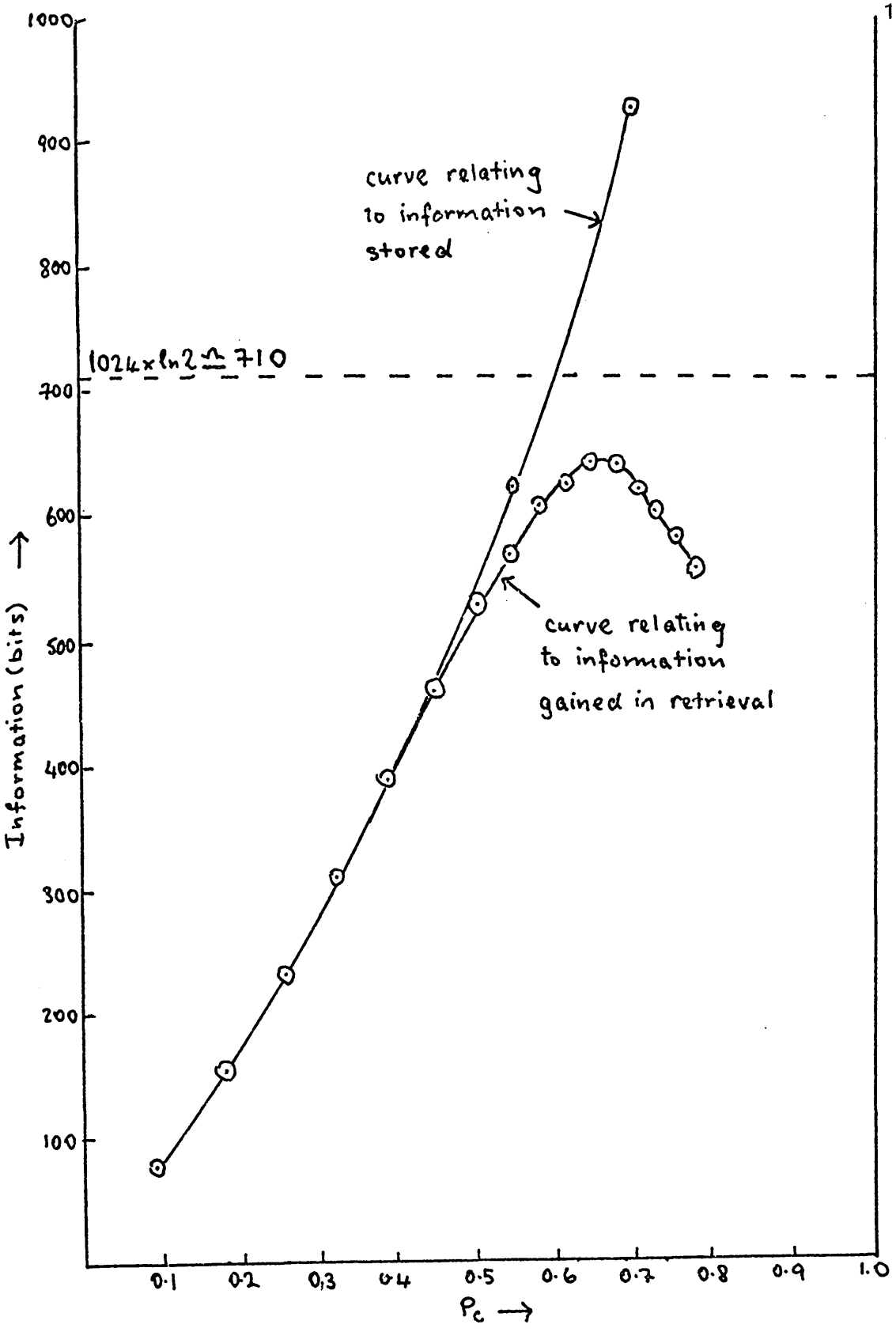

FIGURE 17B. $N=1024$, $M=10$

$NP_c^M = 1$ at $P_c = 0.5$.

CHAPTER 6

The Associative Net. 3

6.1 Introduction

We look at three additional properties of the Associative Net - the structurally simpler non-holographic memory model of the two that we have discussed. We consider a mechanism by means of which the Net may be possibly used indefinitely, we see how it can perform as a recognition device and finally look at the properties it has when it is equipped with a feedback loop.

6.2 Learning with Forgetting

The Associative Net (and the Correlograph) suffers from the disadvantage that only a finite number of pairs of messages may be stored. As long as this number is not exceeded stored messages can be retrieved accurately, but as soon as the store becomes overloaded, retrieval of all messages is severely affected.

The object now will be to show how an Associative Net may be used indefinitely. This is accomplished by means of a mechanism which forgets old associations as new ones are stored.

Consider a square Associative Net (with $M_A=M_B=M$ and $N_A=N_B=N$) loaded to its limit. If a fraction P_c of the nodes has been turned on, then the number of nodes on is P_cN^2.

Storing / . .

Storing a new pair of messages will require $(1-P_c)M^2$ nodes to be turned on since, of the M^2 nodes activated in the Net, P_cM^2 of these are on already. Since increasing the value of P_c may drastically affect the Net's performance, it is required to keep the number of nodes turned on at a constant value.

Let each of the $(1-P_c)M^2$ activated nodes not already on be turned on with probability s. Thus $s(1-P_c)M^2$ of the P_cN^2 nodes already on must be turned off. This is accomplished if each is turned off with probability r where $rP_cN^2 = s(1-P_c)M^2$ 6.1

Consider one node which was activated in the storage of a particular pair of messages and may have been turned on. We shall call this a genuine node. At some stage in the history of the Net, let the probability that it is on be P. After one more recording of another pair of messages the probability P' that it is now on is the probability that it was on before recording and was not turned off plus the probability that it was off and was then turned on, being one of the M^2 activated nodes. This is

$$P' = P(1-r) + \frac{sM^2}{N^2}(1-P) \quad . \ . \ . \ . \ . \ . \ . \ . \ 6.2$$

Writing $Z' = P' - P_c$ and $Z = P - P_c$, then, using equation 6.1, we substitute for P', P and r in 6.2

$P_c + Z' =$ / . .

$$P_c+Z' = \left(1 - \left(\frac{1-P_c}{P_c}\right)s\frac{M^2}{N^2}\right)(P_c+Z) + \frac{sM^2}{N^2}(1-P_c-Z)$$

This simplifies to

$$Z' = Z\left(1 - \frac{sM^2}{P_cN^2}\right)$$

Thus, n associations after the storage of this particular pair of messages, the value of Z has changed from Z(0) to Z(n) where

$$Z(n) = Z(0)\left(1 - \frac{sM^2}{P_cN^2}\right)^n$$

or $Z(n) = Z(0)\exp\left(-\frac{sM^2n}{P_cN^2}\right)$, provided n is not too small and $sM^2 \ll P_cN^2$.

Noting that the number of pairs of messages R stored in order to initially fill up the Net to a density P_c is

$$R = -\frac{N^2}{M^2}\ln(1-P_c) \quad \ldots\ldots\ldots 6.3$$

then $$Z(n) = Z(0)\exp\left(-\frac{sn}{R'}\right) \quad \ldots\ldots 6.4$$

where $$R' = \frac{P_cN^2}{M^2} \quad \ldots\ldots\ldots 6.5$$

R' is approximately equal to R if P_c is small.

When this particular association is first recorded the probability that this genuine node is turned on is

$$P(0) = (1-P_c)s + P_c(1-r) \ .$$

With $$r = s\frac{(1-P_c)}{P_c}\frac{M^2}{N^2} \ ,$$

$$P(0) = / \ . \ .$$

$$P(0) = P_c + s(1-P_c) - s(1-P_c)\frac{M^2}{N^2}$$

$$= P_c + s(1-P_c)(1-\frac{M^2}{N^2}).$$

and since $Z(0) = P(0) - P_c$,

$$Z(0) = s(1-P_c)(1-\frac{M^2}{N^2}) \quad . \quad . \quad . \quad . \quad . \quad . \quad 6.6$$

The probability that after n associations of other pairs of messages this genuine node is on is thus

$$P(n) = P_c + Z(n)$$

$$= P_c + s(1-P_c)(1-\frac{M^2}{N^2})\exp(\frac{-sn}{R'}), \text{ with}$$

$$R' = \frac{P_c N^2}{M^2} \text{(from 6.4 and 6.6)}$$

Here is the problem similar to that of an Associative Net with damaged store (section 5.3). In retrieval, a node which should not transmit a pulse down an output line does so with probability P_c, a node which should transmit a pulse does so with a higher probability

$$P_c + s(1-P_c)(1-\frac{M^2}{N^2})\exp(\frac{-sn}{R'}).$$

The threshold on the output must thus be chosen so that not too many spurious lines output pulses and at the same time not too many genuine lines do not output a pulse.

We consider the case when the threshold is set at M and / . .

and recall must be as good as possible when retrieval of a message just stored is attempted before any other associations have been recorded (i.e. s is set equal to 1). Then, as before, each spurious line will output a pulse with probability P_c^M and, for small error, P_c, M and N are chosen to that $NP_c^M=1$ 6.7

The genuine lines will output a pulse with probability

$$P_{GEN} = \left(P_c + s(1-P_c)(1- \frac{M^2}{N^2})\exp(\frac{-sn}{R'})\right)^M$$

Setting s=1, neglecting $\frac{M^2}{N^2}$ compared to 1 and assuming that $n \ll R'$, then

$$P_{GEN} = \left(P_c + (1-P_c)(1- \frac{n}{R'})\right)^M$$

$$= \left(1 - (1-P_c) \frac{n}{R'}\right)^M$$

$$= 1 - (1-P_c)\frac{Mn}{R'}$$

Allowing that on average one genuine pulse may be missed per message, then this occurs after n associations where n is given by

$$M(1-P_{GEN}) = \frac{M^2}{R'} n(1-P_c) = 1.$$

i.e. $$\frac{n}{R'} = \frac{1}{M^2(1-P_c)}$$

We call n the survival time of a stored message.

Now $R' = \frac{N^2}{M^2}P_c$, and $M = - \frac{\ln N}{\ln P_c}$ (from 6.5 and 6.7)

Consequently / . .

Consequently the survival time of a message is

$$n = \frac{N^2 P_c}{M^4 (1-P_c)}$$

$$n = \frac{(\ln P_c)^4 P_c}{1-P_c} \cdot \frac{N^2}{(\ln N)^4} \quad \ldots\ldots\ldots 6.8$$

Furthermore, from 6.3

$$R = -\frac{N^2}{M^2} \ln(1-P_c) = -\frac{N^2 (\ln P_c)^2 \ln(1-P_c)}{(\ln N)^2} \quad \ldots 6.9$$

With $N = 10^6$ as before, sample values of R and n are:

$P_c = 0.5$, $n = 6.4\ 10^6$, $R = 1.7\ 10^9$

$P_c = 0.3$, $n = 2.5\ 10^7$, $R = 2.7\ 10^9$

$P_c = 0.1$, $n = 8.6\ 10^7$, $R = 2.9\ 10^9$

$P_c = 0.05$, $n = 1.2\ 10^8$, $R = 2.4\ 10^9$

We see that at fixed N the survival time of a message can be lengthened at the expense of loading the Net lightly. We are not able to say whether n can be made the same order as R, although this can be done by substituting an appropriate value of P_c in equations 6.8 and 6.9, since, in fact, one of the assumptions of our analysis was that n should be much less than R.

Computer Simulation Experiments

Values of n, R and M calculated from equations 6.8, 6.9 and 6.7 which are suitable for comparison with simulation results are shown below in Table 5. N was given the value 64

Table 5 / . .

P_c	n	R	M
0.125	37	137	2
0.25	17	131	3
0.354	9	112	4
0.436	5	93	5
0.5	3	79	6
0.561	2	65	7
0.594	1	58	8

Table 5 Values of n,R and M are rounded off to the nearest integer.

The values of N and P_c are required to be set so that the quantities n and M must be appreciably greater than 1. n must be greater than 1 in order that the survival of a message be noticed as further associations are made,and M must be similarly set so that allowing the absence of one genuine pulse in a retrieved message (equation 6.7) (as well as allowing the presence of an extra pulse) implies that the messages are recalled reasonably accurately.

The best that could be done in the simulation, which was an exploratory rather than an exhaustive study, was to choose N=64 and M=6 and initially load the Net half-full./ . .

half-full. Consequently 79 pairs of messages, each message containing 6 components of value 1 generated pseudo-randomly, were stored in a 64x64 Net. 10 pairs of messages were then added one by one to the store by the procedure outlined above, and each member of the pair used to retrieve the other. It was found that the mean lifetime of the first 5 extra pairs added had a value of 2.5 . It was not possible to substantially increase the lifetime above this figure by judicious choice of M and P_c.

Conclusion

Although it has been demonstrated that by means of this mechanism the Associative Net is converted into a relatively short-term memory it has yet to be shown whether the message lifetime can be made substantially large.

6.3 The Associative Net as a recognition device

An Associative Net containing R pairs of messages (A,B) may be required to be used to determine whether a particular pair of messages presented has been previously stored. We cannot, as we can in a local store, search through the individual store locations to look for a match between the message presented and the message in store, since we do not know which store locations (nodes) correspond to which stored messages.

We consider an $N_A x N_B$ Associative Net and restrict the / . .

the possible pairs of messages which may be stored to pairs made up of A-messages containing M_A ones and B-messages containing M_B ones. The procedure is to input one member (say the B-message) of the test pair to the system and compare the output with the test A-message. The pair under consideration is recognised as being stored if all the ones in the test A-message are present in the retrieved A-message. The threshold is set at M_B.

For an $N_A x N_B$ Net, the number of different possible pairs of messages

$$= \binom{N_A}{M_A}\binom{N_B}{M_B}$$

All of the R pairs which have been stored will be recognised. The probability that a particular pair of messages will be recognised as being stored in the Net when it was not is the probability that a particular M_A of the N_A output lines of the Net fire when they should not. This probability is

$$P(\text{error}) = P_c^{M_B M_A}$$

Summed over the ensemble of possible pairs, the mean number of message pairs wrongly classified is thus

$$\left(\binom{N_A}{M_A}\binom{N_B}{M_B} - R\right) P_c^{M_B M_A}$$

The Net will behave almost perfectly if this quantity is made less than 1 and the upper limit on the value of P_c is determined, in the same way as before, by the equality

$$\binom{N_A}{M_A} / \ . \ .$$

$$\binom{N_A}{M_A}\binom{N_B}{M_B} P_c^{M_A M_B} = 1 \quad \ldots\ldots\ldots\ldots 6.10$$

The information gained in placing each of the $\binom{N_A}{M_A}\binom{N_B}{M_B}$ equiprobable pairs in the correct category is

$$I = -\binom{N_A}{M_A}\binom{N_B}{M_B}\left(\frac{R}{\binom{N_A}{M_A}\binom{N_B}{M_B}} \cdot \log_2\left(\frac{R}{\binom{N_A}{M_A}\binom{N_B}{M_B}}\right)\right.$$

$$\left. + \left(1 - \frac{R}{\binom{N_A}{M_A}\binom{N_B}{M_B}}\right) \log_2 \left(1 - \frac{R}{\binom{N_A}{M_A}\binom{N_B}{M_B}}\right)\right) \quad \text{bits}$$

$$\simeq -R\log_2 \left(\frac{R}{\binom{N_A}{M_A}\binom{N_B}{M_B}}\right) \text{ bits}$$

$$\simeq R\ln \left(\binom{N_A}{M_A}\binom{N_B}{M_B}\right) \text{ n.u. provided } R \ll \binom{N_A}{M_A}\binom{N_B}{M_B} \quad . \quad 6.11$$

From 6.10, $\ln \left(\binom{N_A}{M_A}\binom{N_B}{M_B}\right) = - M_A M_B \ln P_c$

and since $R = - \frac{N_A N_B}{M_A M_B} \ln(1-P_c)$,

$$I = N_A N_B \ln P_c \ln(1-P_c) \quad \text{n.u.}$$

As before, the information efficiency $E = I/N_A N_B$ has a maximum at $P_c = \frac{1}{2}$, when $E = \ln 2 \simeq 0.69$

If our condition for good recall holds then, with $P_c = \frac{1}{2}$, certainly

$$\log_2\left(\frac{N_A}{M_A}\right) + \log_2\left(\frac{N_B}{M_B}\right) - M_A M_B \ll 0$$

i.e. / . .

$$\text{i.e. } M_A M_B \gtrsim M_A \log_2 \frac{N_A}{M_A} + M_B \log_2 \frac{N_B}{M_B}$$

This is in fact true if

$$M_A M_B \gtrsim M_A \log_2 N_A + M_B \log_2 N_B$$

or if

$$1 = \frac{\log_2 N_A}{M_B} + \frac{\log_2 N_B}{M_A}$$

This condition is not sufficient to determine both M_A and M_B. One solution would be to put $M_B = 2\log_2 N_A$ and $M_A = 2\log_2 N_B$. In the case of a square Net ($N_A = N_B = N$) where $M_A = M_B = M$ then certainly M is determined, having a value of $2 \log_2 N$. The value of R, is as usual, determined from the equation

$$R = \frac{-N_A N_B}{M_A M_B} \ln(1-P_c).$$

Since analysis has shown that if the Net is to be used efficiently, M_B is to be of the order of $\log_2 N_A$ and M_A is to be of the order of $\log_2 N_B$, even for relatively small values of N_A and N_B (say $N_A, N_B > 30$), the quantities $\binom{N_A}{M_A}\binom{N_B}{M_B}$ are very large compared to R. The approximation used in deriving 6.11 is thus a very good one. Furthermore, under these conditions, only one message out of $\binom{N_A}{M_A}\binom{N_B}{M_B}$ is wrongly classified (equation 6.10). Therefore the conditions of efficient performance obtained may be applied strictly to all but the very smallest size of Associative Net.

Conclusion / . .

Conclusion

The Net does, therefore, work efficiently as a recognition device. As may be expected from previous results, it works best when it is half full but there is some freedom in the choice of the values of M_A and M_B.

6.4 The Associative Net equipped with a feedback loop

Two situations will be considered in the investigation of what the Net can do if it is provided with a feedback loop. This enables the output from the Net to form the input to it for the next retrieval task.

Improving the retrieval of Auto-associations

The first case is when the Net is storing Auto-associations i.e. it stores messages of the form (A,A), and is required to retrieve the whole of a message, given a portion of it. By continually reinputting the output of the Net it firstly emits different messages on successive trials, but eventually it settles down to produce a repeating sequence of messages. The question we shall ask is whether the Net performs any better with this feedback mechanism than without.

Computer simulations were performed on a square Net with $N_A=N_B=N=64$. From the theory previously derived in section 5.5, optimum conditions are attained with $M=2\log_2 N=12$, $P_c=\frac{1}{2}$, $R=\frac{N^2}{M^2}\ln 2 \simeq 20$. In these circumstances, half of a stored message when input will produce as output the whole of the message with small error / . .

error.

20 messages, each containing 12 components out of 64 of value 1, generated pseudo-randomly, were stored in the Net. Fractions of each message were input and the output fed back to the Net until a repeating sequence of messages was produced. In most cases the same message was output successively after a few trials. The performance of the Net with and without feedback was compared by examining the first message output and the final message or set of messages output, and comparing them with the stored message. (Table 6)

The simulation results showed that, in the recall of a message with the Net optimally loaded, there is no difference between the number of errors in the first message output and the number of errors produced in the final output or outputs after cycling. However, cycling does improve the retrieval of some messages at the expense of others for, when the Net is used at its limit, the mean number of messages retrieved perfectly is greater when feedback is used than when it is not.

Table 6 / . .

Cue length as a fraction of the length of a stored message.	Number of messages correctly retrieved		Total number of errors made in the retrieval of 20 messages.	
	With feedback	Without feedback	With feedback	Without feedback
3/12	2	1	97	93
4/12	6	3	58	64
5/12	10	6	31	29
6/12	14	9	16	17
7/12	14	12	11	11
8/12	18	16	6	5

Table 6 M=12, R=20, N=64

Storage of a sequence of messages

A square Associative Net is able to store a complex piece of information which comprises a sequence of R messages by associating the first with the second, the second with the third, the third with the fourth and so on, finally associating the Rth message with the first one. This task is relevant to the problem of how the Net may be used to produce a sequence of simple learned responses which an organism may employ in carrying out a complex task, where the instruction for one particular response / . .

response provides also the information to retrieve the next instruction in the sequence. The whole sequence may be retrieved by feeding in to the Net any one of the R messages in store and reinputting the resultant output. As long as the Net is not overloaded, then the whole sequence of R messages will be retrieved from it. The performance of the Net functioning under these conditions has been investigated by computer simulation.

In the experiments, $N=N_A=N_B$ was given the value of 64. The calculations described earlier (section 4.4) have shown that if one error per message retrieved is allowed, then 79 messages, each having 6 components of value 1, can be stored. In this task, errors are accumulative, as the output in one retrieval is used as the next input to the Net. With the parameters of M and R so set, the sequence of outputs obtained does not resemble the stored sequence. This was found in the simulation. However, by sacrificing information efficiency in increasing the value of M to 10 and decreasing the value of R to 20, stable conditions were obtained. The information efficiency for the retrieval in this way of R-1 stored messages from 1 is

$$- \frac{864.19}{(64)^2}\left(\log_2\left(\frac{10}{64}\right) + \log_2\left(1-\frac{10}{64}\right)\right)$$

$$= \frac{19}{4096}\left(641\log_2 64 - 101\log_2 10 - 541\log_2 54\right)$$

$$= 18\% \ / \ .\ .$$

$= 18\%$.

We check that at this loading the Net is still functioning as a distributed memory. If it were a local memory, both inputs to a particular node would be simultaneously active in the storage of just one message pair. The probability that a particular node is on is

$$P_c = 1 - \left((1-(\frac{10}{64})^2\right)^{20}$$

$$= 1 - \exp\left(\frac{-2000}{4096}\right)$$

$$\simeq 0.39.$$

The probability that a node has both its input lines activated by just one pair of messages to be stored is

$$P_1 = \binom{20}{1}(\frac{10}{64})^2\left((1-(\frac{10}{64})^2\right)^{19}$$

$$\simeq \frac{2000}{4096}.\exp(\frac{-1900}{4096})$$

$$\simeq 0.12$$

Thus approximately 2/3 of the nodes on are identified with more than one pair of stored messages and information is stored in a distributed fashion.

The procedure adopted was to store in the manner just described 20 messages. 10 components out of the 64 of each message were chosen to have the value 1 by means of a pseudo-random number generator. One of these messages was input to the Net. The threshold placed on the / . .

the resulting output, which was initially set equal to the number of ones in the input message, was gradually lowered in steps of one until the number of ones (i.e. pulses) in the output was equal to or greater than the number of ones that the retrieved message should contain- namely 10. This output was then fed in again and the process repeated. On occasions the output differed slightly from what it should have been, but these discrepancies did not propagate as the cycling process proceeded. Although information efficiency has had to be sacrificed to obtain correct regeneration of messages, the advantage is that distorted versions of the stored messages when used as inputs still evoked the correct chain of responses. About 6 out of the 10 components of a previously stored message which have the value 1 could be changed to have the value 0 before the correct cycle of outputs was no longer obtained. At the other extreme, it was safe to change up to about 4 components of value 0 to have the value 1. Table 7 (page 131) shows figures for the allowable amount of distortion to the cue for the intermediate situations when both types of distortion were allowed. In fact, (except in one case) all the messages within a Hamming distance* 4 of a stored message produced the correct cycle of outputs. Assuming that the stored messages are widely separated in N space, this / . .

*The Hamming distance between two binary N-vectors is the number of components in which they differ in value.

this is equivalent to saying that there are about $20\binom{64}{4}$ - say 10^8 - messages which may be used as inputs to evoke the correct cycle.

Since the Net has only a finite number of possible inputs and outputs and one input always produces the same output, then the sequence of outputs produced from any arbitrary input ultimately forms a loop which may contain up to about 2^{64} messages (not exactly 2^{64} as no messages containing less than 10 ones are accepted as outputs). In this case, from the experiments performed, it was found that, for a cue which differs greatly from a stored message (i.e. the Hamming distance between it and each stored message was greater than 4), then the cycle of messages ultimately produced was either that of the loop of stored messages or one of a very few other loops of length no greater than about 100.

Table 7 / . .

G ↓	S → 0	1	2	3	4
0	*	*	*	*	x
1	*	*	*	*	
2	*	*	*		
3	*	*			
4	*	*			
5	*	*			
6	*				
7					
8					

Table 7 N=64, M=10, R=20.

Distorted versions of 6 of the 20 stored messages were used as inputs. The distortion is characterised by the number G - the number of components of such a message of value 1 which are assigned the value 0 and S - the number of components of value 0 which are given the value 1. A star (*) in table 7 signifies that all 6 messages with that pair of values of G and S produced ultimately the correct cycle of outputs. For G=0 and S=4, only one of the 6 messages so distorted failed to produce the correct outputs. This is signified by a cross (x).

For / . .

For other values of G and S, the sequence of outputs did not stabilise correctly. Two simulations were looked at in detail. In the first case, of 30 arbitrarily chosen inputs, 20 ultimately produced the same cycle of length 50 and the other 10 produced the stored cycle. In the second case, 9 arbitrarily chosen input messages produced the stored cycle, (of length 20) which was the only cycle ever found.

Conclusions

There are stable conditions under which the N*N Associative Net may store a chain of messages, Although these are attained at the expense of saarificing information efficiency, there is the compensating factor that the system as a consequence performs very well when given damaged cues. From the small number of simulation experiments performed, the indications are that the large number (approximately 2^N) of possible inputs produce a very small number of relatively short output cycles.

CHAPTER 7

Symmetrical Associative Nets

7.1 Introduction

The Associative Net store is essentially an array of binary switching elements. In the storage process, although a switch registers the presence of activity in both of the lines to which it is connected, it does not distinguish between the other situations of activity in just one line and not the other and no activity in either. We now consider models similar to the Associative Net, but which were expressly designed to use this discarded information, and look at their information efficiency.

7.2 The Symmetrical Associative Net

Like the Associative Net this comprises a set of N_A parallel lines which cross another set of N_B parallel lines at right angles. Information about pairs of messages is stored by sending pulses down both the A-lines and the B-lines and activating the memory elements which are situated at the $N_A N_B$ intersections of the sets of parallel lines. We call these elements "switches" to distinguish them from the "nodes" of an Associative Net. We shall represent the messages input to the Net as vectors whose components may take the values +1 or -1 (instead of the values 1 or 0 as for the Associative Net). Each switch counts coincidences and anti-coincidences. For / . .

For example, if positive pulses are sent along lines A_1 and B_2 and negative pulses are sent along lines A_2 and B_1, then switches (A_1,B_2) and (A_2,B_1) will each record a coincidence and (A_1,B_1) and (A_2,B_2) an anti-coincidence. As the messages are stored, count is kept of the difference between the number of coincidences and the number of anti-coincidences at each switch. With the N_AxR matrix A and the N_BxR matrix B containing information about the R stored message pairs $(A^{(r)},B^{(r)})$ where r=1,2,..R, this is accomplished if the sum

$$C_{ij} = \sum_{r=1}^{R} A_{ir}B_{jr}$$

is computed for all allowable values of i and j. When all messages have been stored, each switch (A_i,B_j) is given the value +1 or -1 according to the sign of the sum C_{ij}. (An odd number of messages is stored so that each switch does have a definite value).

In the retrieval process, if say message $A^{(n)}$ is required, the B-lines are activated by the pulses of message $B^{(n)}$. Consider the output down line A_i. This passes through N_B switches. Each of these which has value +1 will output a pulse whose sign is the same as the one it receives. Conversely, each of these which has value -1 outputs a pulse of opposite sign to the one input. These pulses feed into a threshold device which emits a positive or a negative pulse depending on the number of positive and negative pulses it receives. This final / . .

final output α_i is compared with the component A_{in} of the original stored message and can be expressed as

$$\alpha_i = \left[\sum_s C_{is} B_{sn} \right] \quad \text{where} \quad C_{ij} = \left[\sum_{r=1}^{R} A_{ir} B_{jr} \right]$$

Equations 4.7 and 4.8 (page 73A) show that the relation between this model and the Associative Net is not merely structural. In particular, the Symmetrical Associative Net is also a <u>non-linear</u> model in the strict sense of section 1.6.

We shall consider how the memory functions in storing and retrieving R message pairs. N_A, N_B and R are all odd numbers as majority votes are taken, and we shall introduce the quantities n_a, n_b and ρ, where

$$N_A = 2n_a+1, \quad N_B = 2n_b+1, \quad R = 2\rho+1$$

Having stored the R message pairs, we now wish to recover one of the A-messages $A^{(n)}$. Consequently, the appropriate B-message $B^{(n)}$ is input along the B-lines. Looking at output line A_i, for example, this has N_B pulses travelling down it. One of these was transmitted by switch (A_i, B_j) and will have the same sign as the pulse A_{in} if this switch has the same parity as the pair (A_{in}, B_{jn}) (It has even parity if $A_{in}B_{jn}=1$ and odd parity if $A_{in}B_{jn}= -1$)

This will be so if at least ρ of the 2ρ other pairs of messages stored have the same parity as (A_{in}, B_{jn}).

Let the B-messages have as components 1's and -1's chosen / . .

chosen with equal probability. Then the probability that switch (A_i, B_j) sends the correct pulse down line A_i is

$$P_{switch} = \binom{2p}{p}\left(\tfrac{1}{2}\right)^{2p} + \binom{2p}{p+1}\left(\tfrac{1}{2}\right)^{2p} + \binom{2p}{p+2}\left(\tfrac{1}{2}\right)^{2p} + \ldots + \binom{2p}{2p}\left(\tfrac{1}{2}\right)^{2p} \quad \ldots\ldots 7.1$$

This will be written as

$$P_{switch} = P_0 + P_1 \quad \ldots\ldots\ldots\ldots\ldots 7.2$$

where $P_0 = \binom{2p}{p}\left(\tfrac{1}{2}\right)^{2p}$ $\ldots\ldots\ldots\ldots\ldots 7.3$

If p is large then we may use the more precise form of Stirling's approximation - namely that

$$\ln(p!) = (p+\tfrac{1}{2})\ln p - p + \tfrac{1}{2}\ln 2\pi.$$

Hence from 7.3

$$\begin{aligned} \ln P_0 &= \ln(2p!) - 2\ln(p!) \\ &\simeq (2p+\tfrac{1}{2})\ln(2p) - 2(p+\tfrac{1}{2})\ln 2 - \tfrac{1}{2}\ln 2\pi - 2p\ln 2 \\ &= -\tfrac{1}{2}\ln(p) - \tfrac{1}{2}\ln\pi \end{aligned}$$

$$\text{i.e. } P_0 \simeq (\pi p)^{-\frac{1}{2}} \quad \ldots\ldots\ldots\ldots\ldots 7.4$$

Now P_0 is the central term of a symmetrical binomial distribution whose sum is 1.

$$\text{i.e. } 1 = \sum_{k=0}^{2p}\binom{2p}{k}\left(\tfrac{1}{2}\right)^{2p} = \sum_{k=0}^{p-1}\binom{2p}{k}\left(\tfrac{1}{2}\right)^{2p} + \binom{2p}{p}\left(\tfrac{1}{2}\right)^{2p} + \sum_{k=p+1}^{2p}\binom{2p}{k}\left(\tfrac{1}{2}\right)^{2p}$$

$$= P_0 + 2P_1 \quad \ldots\ldots\ldots 7.5$$

(from 7.1 and 7.3)

From / . .

From 7.2, 7.4 and 7.5,

$$P_{switch} = P_0 + P_1$$
$$= P_0 + \frac{1-P_0}{2}$$
$$= \tfrac{1}{2}(1+P_0) \quad \ldots\ldots\ldots 7.6$$
$$\simeq \tfrac{1}{2}(1+(\pi\rho)^{-\frac{1}{2}})$$

which is the probability that one of the N_B switches linked by line A_i sends the correct pulse down it. P_{switch} decreases to a limit of 0.5 as the number R of messages stored increases ($R=2\rho+1$).

In order to determine the final output from line A_i, a majority vote is taken of the N_B pulses travelling down it. An incorrect decision is made if n_b+1 or more of the $2n_b+1$ switches of that line have transmitted the incorrect pulse.

The probability of a mistake is thus

$$P_{mistake} = \sum_{i=1}^{n_b+1} \binom{2n_b+1}{n_b+i} (1-P_{switch})^{n_b+i} (P_{switch})^{n_b+1-i}$$

From 7.6, substituting for P_{switch} in terms of P_0,

$$P_{mistake} = \sum_{i=1}^{n_b+1} \binom{2n_b+1}{n_b+i} (\tfrac{1}{2})^{2n_b+1} (1-P_0)^{n_b+i} (1+P_0)^{n_b+1-i} \quad .7.7$$

Consider the first term Q_1 of this sum

$$Q_1 = \binom{2n_b+1}{n_b+1} (\tfrac{1}{2})^{2n_b+1} (1-P_0)^{n_b+1} (1+P_0)^{n_b} \quad \ldots 7.8$$

Using Stirling's approximation once more,

ln / . .

$$\ln\left(\binom{2n_b+1}{n_b+1}\left(\tfrac{1}{2}\right)^{2n_b+1}\right) \simeq (2n_b+\tfrac{3}{2})\ln(2n_b+1) - (n_b+\tfrac{3}{2})\ln(n_b+1) - (n_b+\tfrac{1}{2})\ln(n_b) - (2n_b+1)\ln 2 - \tfrac{1}{2}\ln 2\pi$$

For n_b large $\ln(n_b+1) \simeq \ln(n_b)$ and $\ln(2n_b+1) \simeq \ln(2n_b)$

Consequently

$$\ln\left(\binom{2n_b+1}{n_b+1}\left(\tfrac{1}{2}\right)^{2n_b+1}\right) \simeq -\tfrac{1}{2}\ln(\pi n_b)$$

i.e. $$\binom{2n_b+1}{n_b+1}\left(\tfrac{1}{2}\right)^{2n_b+1} \simeq (\pi n_b)^{-\frac{1}{2}}$$

and substituting in 7.8, $$Q_1 \simeq \frac{(1-P_0)^{n_b+1}(1+P_0)^{n_b}}{(n\pi)^{\frac{1}{2}}} \quad . \; . \; 7.9$$

Now Q_1 is the first term of a series whose sum is the probability of error in one component of the output (equation 7.7). This is a difficult series to sum exactly, but a good upper limit can be obtained.

The ith and the (i+1)th terms of this series are respectively

$$Q_i = \binom{2n_b+1}{n_b+i}\left(\tfrac{1}{2}\right)^{2n_b+1}(1-P_0)^{n_b+i}(1+P_0)^{n_b+1-i},$$

$$Q_{i+1} = \binom{2n_b+1}{n_b+i+1}\left(\tfrac{1}{2}\right)^{2n_b+1}(1-P_0)^{n_b+i+1}(1+P_0)^{n_b+1-(i+1)}$$

The ratio of successive items is thus

$$\frac{Q_{i+1}}{Q_i} = \frac{(2n_b+1)!(n_b+i)!(n_b+1-i)!(1-P_0)^{n_b+i+1}(1+P_0)^{n_b+1-(i+1)}}{(2n_b+1)!(n_b+1+i)!(n_b-i)!(1-P_0)^{n_b+i}(1+P_0)^{n_b+1-i}}$$

$$= / \; . \; .$$

$$= \frac{(n_b+1-i)(1-P_0)}{(n_b+1+i)(1+P_0)} \quad \ldots\ldots\ldots\ldots 7.10$$

$$\leqslant \frac{1-P_0}{1+P_0} \qquad \left(\frac{n_b+1-i}{n_b+1+i} < 1 \text{ since } i>0\right)$$

and so $Q_i \leqslant Q_1 \left(\frac{1-P_0}{1+P_0}\right)^{i-1}$

Consequently 7.7. becomes

$$P_{\text{mistake}} \leqslant Q_1 \sum_{i=1}^{n_b+1} \left(\frac{1-P_0}{1+P_0}\right)^{i-1}$$

$$= \frac{Q_1\left(1-\left(\frac{1-P_0}{1+P_0}\right)^{n_b}\right.}{1-\frac{1-P_0}{1+P_0}}$$

$$= \frac{Q_1}{2}\left(1+\frac{1}{P_0}\right)\left(1-\left(\frac{1-P_0}{1+P_0}\right)^{n_b}\right) \quad \ldots\ldots 7.11$$

Substituting for Q_1 from 7.9,

$$P_{\text{mistake}} \leqslant \frac{(1-P_0)^{n_b+1}(1+P_0)^{n_b}}{2(n\pi)^{\frac{1}{2}}}\left(1+\frac{1}{P_0}\right)\left(1-\left(\frac{1-P_0}{1+P_0}\right)^{n_b}\right)$$

$$= \frac{(1-P_0^2)^{n_b+1}}{2(n\pi)^{\frac{1}{2}}P_0}\cdot\left(1-\left(\frac{1-P_0}{1+P_0}\right)^{n_b}\right)$$

Now $\frac{1-P_0}{1+P_0} \doteq 1-2P_0$ for $P_0 \ll 1$.

and if n_b is large, by using the fact that

$$\underset{t\to\infty}{\text{Lt}} \left(1-\frac{x}{t}\right)^t = \exp(-x) \quad ,$$

then / . .

then

$$P_{mistake} \ll \frac{1}{2P_0(n\pi)^{\frac{1}{2}}} \exp(-P_0^2(n_b+1))(1-\exp(-2n_bP_0)) \quad .7.12$$

$$P_{mistake} \ll \tfrac{1}{2}\left(\frac{\rho}{n_b}\right)^{\frac{1}{2}} \exp\left(-\frac{n_b+1}{\pi\rho}\right)\left(1-\exp\left(\frac{-2n_b}{(\pi\rho)^{\frac{1}{2}}}\right)\right) \text{(using 7.4)}$$

$$\simeq \tfrac{1}{2}\left(\frac{\rho}{n_b}\right)^{\frac{1}{2}} \exp\left(-\frac{n_b}{\pi\rho}\right)\left(1-\exp\left(\frac{-2n_b}{(\pi\rho)^{\frac{1}{2}}}\right)\right)$$

Let the A-messages have as their components 1's and -1's chosen with equal probability. Then the amount of information stored in the Net, which is subsequently retrieved with a probability of error given by $P_{mistake}$, is RN_A bits. If this system were used to its information theoretical limit, N_AN_B bits could be held reliably. A quantity D, which is a measure of how heavily the store is loaded, is defined as the ratio of these two quantities.

$$D = \frac{RN_A}{N_AN_B} = \frac{2\rho+1}{2n_b+1} \simeq \frac{\rho}{n_b} \quad \text{(if } \rho \gg 1 \text{ and } n_b \gg 1)$$

Consequently we can write

$$P_{mistake} \ll \tfrac{1}{2}(D)^{\frac{1}{2}} \exp\left(\frac{-1}{\pi D}\right)\left(1-\exp\left(-\frac{2}{D}\left(\frac{\rho}{\pi}\right)^{\frac{1}{2}}\right)\right)$$

$$\simeq \tfrac{1}{2}(D)^{\frac{1}{2}} \exp\left(\frac{-1}{\pi D}\right) \text{ for } \rho \gg 1 \text{ and provided D is not greater than 1, say.} \quad \ldots\ldots\ldots 7.13$$

This expression sets an upper limit on the value of $P_{mistake}$. If it is to be regarded as a good approximation to / . .

to $P_{mistake}$ the approximation

$$\frac{n_b-i+1}{n_b+i-1} = 1$$

must be valid (see equation 7.10). It is thus required that the ratio

$$\frac{n_b-i+1}{n_b+i+1}$$

be very much closer to 1 than the ratio $(1-P_0)/(1+P_0)$

$$\text{i.e.} \quad 1 - \frac{n_b-i+1}{n_b+i+1} \ll 1 - \frac{(1-P_0)}{1+P_0}$$

$$\text{or} \quad i \ll P_0(n_b+1) \quad \ldots\ldots\ldots\ldots 7.14$$

This is not satisfied for all i since $P_0 \ll 1$ and i varies from 1 to n_b+1. But if the series

$$\sum_{i=1}^{n_b+1} \left(\frac{1-P_0}{1+P_0}\right)^{i-1}$$

converges well before i reaches the value n_b+1, then the effective range of i is reduced (see equation 7.10).

The above series converges when the value of i is such that

$$\left(\frac{1-P_0}{1+P_0}\right)^{i-1} \ll 1$$

i.e. $(i-1).\ln\left(\frac{1-P_0}{1+P_0}\right)$ should be large and negative.

or $(i-1).-2P_0 \ll -1 \qquad (P_0 \ll 1 \text{ so } \ln(1\pm P_0) \simeq \pm P_0)$

Consequently $i \gg \frac{1}{2P_0} + 1$

Setting / . .

Setting $i_{max} = \frac{2}{P_0}$, so that we assume that the only terms contributing non-neglibly to the sum of the series

$$\sum_{i=1}^{n_b+1} \left(\frac{1-P_0}{1+P_0}\right)^{i-1}$$

are those which have values of i not greater than $\frac{2}{P_0}$, then going back to the original condition on i (7.14),

$$\frac{2}{P_0} \ll P_0(n_b+1) \simeq P_0 n_b .$$

Hence $$P_0^2 \gg \frac{2}{n_b} .$$

With $$P_0 \simeq \frac{1}{(\pi\rho)^{\frac{1}{2}}} \quad (7.4),$$

$$\rho \ll \frac{n_b}{2\pi}$$

i.e. $$\frac{\rho}{n_b} \ll \frac{1}{2\pi}$$

which is satisfied if the packing density $D \simeq \frac{\rho}{n_b}$ is not too large.

If this is so, we can set the probability of error $P_{mistake}$ equal to P_{max} where (equation 7.13)

$$P_{max} = \tfrac{1}{2} D^{\frac{1}{2}} \exp\left(\frac{-1}{\pi D}\right) .$$

Values of P_{max} for various values of D are shown in Table 8

Information retrieval efficiency

Let R pairs, each made up of an N_A-bit and an N_B-bit message, be stored to a density specified by $D=R/N_B$. If all / . .

all A-messages were retrieved perfectly then the information gained would be RN_A bits. There is a probability, however, that a component of each A-message is retrieved with error $P_{mistake}$. The number of bits required to be supplied per component to clear up this error is

$$- (P\log_2 P + (1-P)\log_2(1-P)) \text{ bits}$$

where P is written for $P_{mistake}$, so that the total amount of information gained in retrieving the R messages stored in the Net is

$$I = RN_A(1+P\log_2 P + (1-P)\log_2(1-P)) \text{ bits} \quad . \quad . \quad . \quad . \quad . \quad 7.15$$

The maximum number of bits which can ever be gained from this store, which contains $N_A N_B$ binary registers, is $N_A N_B$ bits. Consequently, the efficiency of retrieval is

$$E = \frac{I}{N_A N_B} = D(1 + P\log_2 P + (1-P)\log_2(1-P))$$

$$\text{with } P \lesssim \tfrac{1}{2}D^{\frac{1}{2}}\exp(\tfrac{-1}{\pi D})$$

With P set at $\frac{1}{2}D^{\frac{1}{2}}(\frac{-1}{\pi D})$, Table 8 also shows values of E expressed as a function of D.

Table 8 / . .

Information storage density D	Efficiency of information retrieved E	Upper bound on the probability that one component of a retrieved message is in error $P = \frac{1}{2}D^{\frac{1}{2}}\exp(\frac{-1}{\pi D})$
0.05	0.0499	0.0002
0.1	0.094	0.007
0.15	0.126	0.023
0.2	0.147	0.046
0.25	0.159	0.070
0.3	0.164	0.095
0.35	0.166	0.119
0.4	0.159	0.143
0.6	0.135	0.228
0.8	0.095	0.300

Table 8

7.3 The effect of damage to the Symmetrical Associative Net

Let there be a probability 1-Q that each of the switches of this Net is damaged - i.e. that the switch has the opposite parity to what it should have. Considering once more the retrieval of a stored A-message by inputting the appropriate B-message, in the absence of damage each of the N_B switches traversed by a particular A-line / . .

A-line transmits a pulse of the correct polarity with probability

$$P_{switch} = \tfrac{1}{2}(1+P_0) \quad \ldots\ldots\ldots 7.16$$

where $P_0 \approx (\pi\rho)^{-\frac{1}{2}}$ (from 7.2 and 7.4)

A damaged switch will, however, transmit a pulse of the correct polarity with probability

$$P_{switch}(Q) = P_{switch}Q + (1-P_{switch})(1-Q).$$

Writing $P_{switch} = \tfrac{1}{2}(1+P_0)$ and $Q = \tfrac{1}{2}(1+q)$, then

$$P_{switch}(Q) = \tfrac{1}{4}(1+P_0)(1+q) + \tfrac{1}{4}(1-P_0)(1-q)$$

$$= \tfrac{1}{2}(1+P_0q) \quad \ldots\ldots\ldots 7.17$$

Comparing 7.16 and 7.17, then the effect of damage is to replace P_0 everywhere by P_0q.

Therefore, from 7.12,

$$P_{mistake}(Q) \lesssim \tfrac{1}{2} \frac{1}{P_0q(\pi n)^{\frac{1}{2}}} \exp(-P_0^2q^2(n_b+1))(1-\exp(-2n_bP_0q))$$

$$\approx \tfrac{1}{2}\frac{(D)^{\frac{1}{2}}}{q} \exp(\frac{-q^2}{\pi D}) \quad \ldots\ldots\ldots 7.18$$

(by analogy to equation 7.13)

- as opposed to the undamaged case

$$P_{mistake} \lesssim \tfrac{1}{2}(D)^{\frac{1}{2}}\exp(\frac{-1}{\pi D}) \quad \ldots\ldots\ldots (7.13)$$

Thus, in the face of damage, the Net can perform equally well if a lower packing density is accepted - i.e. if less messages are stored. For a given error rate, since $D=R/N_B$, the ratio of the number of message pairs stored / . .

stored to the number stored without damage is

$$\frac{R(Q)}{R(1)} = \frac{D(Q)}{D(1)} = q^2 = (2Q-1)^2$$

To compare how the Net can perform in comparison to the best which may ever be expected from a store of $N_A N_B$ binary registers damaged in this way, it should be noted that the maximum amount of information which can ever be gained from such a damaged binary register is

$$1 + Q\log_2 Q + (1-Q)\log_2(1-Q) \text{ bits.}$$

By reference to 7.15 and 7.18, the efficiency of retrieval in this damaged Net is

$$E(Q) = \frac{D\left(1 + P(Q)\log_2 P(Q) + \left(1-P(Q)\right)\log_2\left(1-P(Q)\right)\right)}{1 + Q\log_2 Q + (1-Q)\log_2(1-Q)}$$

where

$$P(Q) = P_{\text{mistake}}(Q) \simeq \tfrac{1}{2}\,\frac{D^{\frac{1}{2}}}{2Q-1}\,\exp\left(\frac{-(2Q-1)^2}{\pi D}\right),$$

with the same assumptions as before.

At fixed Q, the maximum value of E(Q,D) varied as a function of the information storage density D was found numerically for a range of values of the damage parameter Q. These maxima are listed below in Table 9.

Table 9 / . .

Q	max. value of E(Q,D)	D	P_{max}
1	0.166	0.35	0.119
0.9	0.200	0.22	0.116
0.8	0.214	0.12	0.111
0.7	0.222	0.05	0.101
0.6	0.228	0.015	0.013
0.5	0	-	-
0.4	0.228	0.015	0.013
0.3	0.222	0.05	0.101
0.2	0.214	0.12	0.111
0.1	0.200	0.22	0.116
0	0.166	0.35	0.119

Table 9

7.4 Conclusions drawn from the analysis of the Symmetrical Associative Net

Both damaged and undamaged Symmetrical Associative Nets can be made to store and retrieve information about 20% as efficiently as a system made up of similar components working at the information theoretical limit (see Tables 8 and 9). The probability of error in any one component of a retrieved message may be, however, rather high*, with a value of about 0.1. Thus, like a finite size Associative / . .

* P_{max} is an upper limit to $P_{mistake}$.

size Associative Net, the system works best by storing a lot of messages, each of which is not reproduced very faithfully on recall. As in the case of the Associative Net, the probability of error per message retrieved can be reduced at the expense of lowering the information efficiency of the system.

Where this model does fall down is that, once definite parities have been assigned to switches, unless each switch $(i,j)^*$ retains a note of the number of coincidences relative to the number of anti-coincidences (which is done by computing the sum

$$C_{ij} = \sum_{r=1}^{R} A_{ir} B_{jr}),$$

it is impossible for a new pair of messages to be stored in the Net. Furthermore, it would be difficult for an odd number of pairs to be always stored, so that each switch would be assured of having a definite parity. Storing the values of C_{ij} would have the effect of reducing the information efficiency by a factor of $\log_2 R$. This is in payment for knowledge of the relative importance of the switches, which is not used if each switch is regarded as a binary element.

7.5 The Weighted Symmetrical Associative Net

A simple way of employing this information would be to assign to each switch (i,j) a weight

$$C_{ij} = / \ . \ .$$

* "switch(i,j)" is synonymous with "switch (A_i, B_j)"

$$C_{ij} = \sum_{r} A_{ir} B_{jr}.$$

On retrieval of message $A^{(n)}$ by inputting $B^{(n)}$, switch (i,j), for example, sends C_{ij} pulses down line A_i. All these pulses have the same sign - that of $C_{ij}B_{jn}$. Taking into account contributions from all switches which are linked by line A_i, the sign of

$$\sum_{j=1}^{N_B} C_{ij} B_{jn}$$

determines whether the final output of this line is a positive or a negative pulse.

The mathematics derived in section 7.2 no longer applies to this situation. The efficiency of this distributed memory, which will be called the Weighted Symmetric Associative Net (W.S.A.N.) can, however, be discussed and it is here appropriate to refer to the earlier remark that the Associative Net is a non-linear analogue of the Off-diagonal Holophone. The W.S.A.N. is similar to the Associative Net and in fact even more similar to the Off-diagonal Holophone. Using $B^{(n)}$ as an input, the final output from line A_i of the W.S.A.N. is a positive or negative pulse depending on the sign of

$$\alpha_i = \sum_{j=1}^{N_B} C_{ij} B_{jn} = \sum_{j=1}^{N_B} \sum_{r=1}^{R} A_{ir} B_{jr} B_{jn}. \quad . \quad 7.19$$

If / . .

If now the same **message** were input along the A-lines and the B-lines in the storage process and a fraction of a stored **message** were input along the B-lines to retrieve the rest down the A-lines then, apart from the matter of the threshold on the final output, the mathematics of the W.S.A.N. and the Off-diagonal Holophone are identical (see equation **3.23A**, page 50).

In the more general case of each member of a stored pair being different then, on retrieval of message $A^{(n)}$, for the output line A_i, for example, the number α_i (equation 7.19) is drawn from a normal distribution of mean $\mu = \pm N_B$ and variance $\sigma^2 = (R-1)N_B$. The sign of μ depends on the sign of the pulse originally sent down the line during storage of $A^{(n)}$. Since positive and negative pulses are treated here on an equal footing, then the threshold must be set at a value of 0. Then the threshold device on line A_i will respond incorrectly if the sign of α_i is different from that of the component A_{in} of the stored message. If the value of N_B is large enough for continuous notation to be adopted, then the probability of error per component of an output **message** is the area under the tail of a normal distribution curve and of value

$$P_{error} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left(-\frac{(|\mu| - x)^2}{2\sigma^2}\right) dx.$$

$$= / \ . \ .$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{-|\mu|/\sigma} \exp\left(-\tfrac{1}{2}y^2\right)dy$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{-(S/N)^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}y^2\right)dy \quad \ldots\ldots\ldots 7.20$$

where the signal to noise ratio S/N is defined as $\frac{\mu^2}{\sigma^2}$ (as in Chapter 3).

By analogy to equation 3.25 (page 51) the signal to noise ratio is

$$S/N = \frac{N_B}{R-1} \quad \ldots\ldots\ldots\ldots 7.21$$

and thus is related to the information storage density

$$D = \frac{R}{N_B \log_2 R} \doteq \frac{1}{S/N \log_2 R} \text{ for } N_B \text{ large} \quad \ldots\ldots 7.22$$

Values of D which must not be exceeded to ensure almost perfect retrieval (one error per retrieved message is allowed) are shown below in Table 10 for different values of N_A, where $N_A = N_B = N$.* For N_A sufficiently large we can assume the efficiency of information retrieved E to be also equal to D. It is not necessary to discuss in detail the implications of allowing one error per retrieved message as we did for the Associative Net (section 5.2) for, by reason of the high information content of the message stored, this criterion here defines very accurate conditions of retrieval. It will be seen that / . .

---

* In these calculations we set $N.P_{error}=1$ and use 7.21 and 7.22.

that the information efficiency is only of the order of a few per cent.

N	signal to noise ratio $S/N = \frac{\mu^2}{\sigma^2}$	$\log_2 R = \log_2(\frac{N}{S/N})$	Information density/efficiency $= \frac{1}{S/N \log_2 R}$
10^2	5.48	4.19	0.043
10^3	9.0	6.80	0.016
10^4	13.69	9.52	0.008
10^5	18.06	12.4	0.005
10^6	23.04	15.4	0.004

Table 10 $N_A=N_B=N$.

For completeness, the signal to noise ratio for a Holophone which has a threshold detector placed on its output will also be related to the probability of error per component of a retrieved message. For this system we have (equation 3.22, page 44) $\mu = \pm N'$ and $\sigma^2 = N'(NR-1)$

i.e. $S/N = \frac{\mu^2}{\sigma^2} = \frac{N'}{(NR-1)}$,

where a message of length N' is used as cue to retrieve the rest of it from a store of R messages each of length N. Unless R equals 1 and N'=N-1, the value of S/N can never / . .

never be greater than 1. In equation 7.20, P_{error}, which increases as S/N decreases, has the value 0.16 for S/N equal to 1 so that, unless only one message is stored and the cue is absurdly long, there is a large chance that a component of the retrieved message is incorrect. This reinforces the conclusions of Chapter 3 that the Holophone is not suited for storing messages of the type considered - i.e. those containing no redundancy.

7.6 Symmetrical Associative Nets - Conclusions

It will be remembered that the Associative Net performs best in storing highly redundant messages; the systems considered here deal with messages containing no redundancy.

The non-linear Symmetrical Associative Net was found to perform relatively efficiently and did well in the face of damage. Although general statements concerning its performance could be made, there are accompanying disadvantages in its lack of versatility.

The less efficienct W.S.A.N., which we treat as a linear device although it does have a threshold device placed on its output, has storage elements which are no longer binary valued. This raises the question as to the relationship between the distributed memory Nets that we have discussed and linear Perceptrons. We will look at related systems, such as the Perceptron, in the next chapter.

CHAPTER 8

Related Systems

8.1 Introduction

The linear Perceptron (or Adaline) does show similarities in structure to the three distributed memory Nets that we have considered - namely the Associative Net, the Symmetrical Associative Net and the W.S.A.N. We look at this relationship and, in line with many Perceptron studies, consider the methods we have employed for modifying the storage elements of the Nets. We discuss a punched card retrieval system which is similar to the Associative Net and then look at the work of Gabor which is closely related to our studies of the Holophone and the Correlograph.

8.2 Perceptrons and Distributed Memory Nets

Farley and Clark (1954) were probably the first to investigate how ensembles of neuron-like elements could be taught to carry out particular specified tasks. They simulated by computer the performance of a set of pseudo-randomly interconnected neuron-like elements. With some of these being designated as input elements and some as output elements, the system was required to learn to discriminate between two periodically time varying patterns of stimulation of the input elements. The learning procedure consisted of increasing the strength

of / . .

of connections between all those elements which took part in a correct discrimination (indicated by particular activity in the output elements) and decreasing the strength of appropriate connections if the discrimination was incorrect. In one set of experiments, of 30 different nets, none having more than 64 elements, all but 4 of them were able to learn to discriminate between a particular pair of input patterns. (We note that Farley and Clark had their eyes on Lashley's findings for in exploratory experiments they found that "arbitrary destruction of at least 10% of the elements may be sustained without impairment of performance".)

A large number of systems of this type, such as the Perceptron in its original form and the Adaline, have been constructed, many of which were presented as models of ensembles of nerve cells although their elements bore little resemblance to what is known about real neurons. Farley and Clark, however, hesitated to claim that their model had any strong ties with neurophysiology.

The general principle behind both the Perceptron, which was invented by Rosenblatt (Rosenblatt 1958, 1962) and has since appeared in many manifestations, and the Adaline (Widrow 1964) can be explained by considering these machines, which are required to learn to perform binary / . .

binary classifications, to be made up of a number of variable gain amplifiers. Each amplifier input is connected to a subset of points of an N point binary retina, while the outputs are linked up in parallel. When a message V, which is a time independent pattern of excitation on the retina, is presented for classification, a pulse of unit height is sent into an amplifier if all the retinal points to which it is connected are active. After passage through the amplifiers, the single summed output P from the system is compared with a number θ, the threshold, and the message V is put into one category if P is greater than θ and into the other otherwise. If we represent the values of the input and the gains of all the amplifiers (which number say M) by the M-dimensional vectors X and W, where the components of X have the value of 0 or 1, then the machine in fact computes the scalar product $P=\underline{X}.\underline{W}$.

The word Perceptron describes a system of this type which might have any pattern of connection between retina and amplifiers. In the Adaline, a specialised Perceptron, each retinal point is connected to one and only one amplifier (so that M=N). It is a linear device in that the sum to be compared with θ is $P=\underline{V}.\underline{W}$, which is a weighted sum of the individual components of the pattern V to be classified. It has been proved that it is able to learn to perform any binary classification of any

number / . .

number of messages provided that, if each message be represented as a point in N-dimensional space, the two categories can be separated by a hyperplane. (see for example Block 1962, Minsky and Papert 1969).

It is not possible to say much more about how Perceptrons learn to classify patterns except that, if all possible wiring schemes are available, then a Perceptron can perform any binary classification but (apart from the linear case) it is not known how this is done.

A memory Net with N_B input lines and N_A output lines is in structure a battery of N_A Adalines or linear Perceptrons (see Figure 18, page 158). Each of these has N_B inputs and a common output. Each input line is connected to an amplifier (the switches or nodes of the Net) and there is a threshold detector placed in the output line. The gains of the amplifiers, whose values will be represented by the components of the N_A N_B-dimensional vectors $W^{(i)}$, are adjusted during storage according to a simple rule. In the retrieval of message $\underline{A}^{(n)}$, message $\underline{B}^{(n)}$ is input to each of the N_A Perceptrons, and the output of strength equal to the scalar product $\underline{B}^{(n)} . \underline{W}^{(i)}$ is fed into the non-linear detector to determine the final output. The value of the threshold θ is common to all N_A Perceptrons.

In function, however, memory Nets can not be viewed as / . .

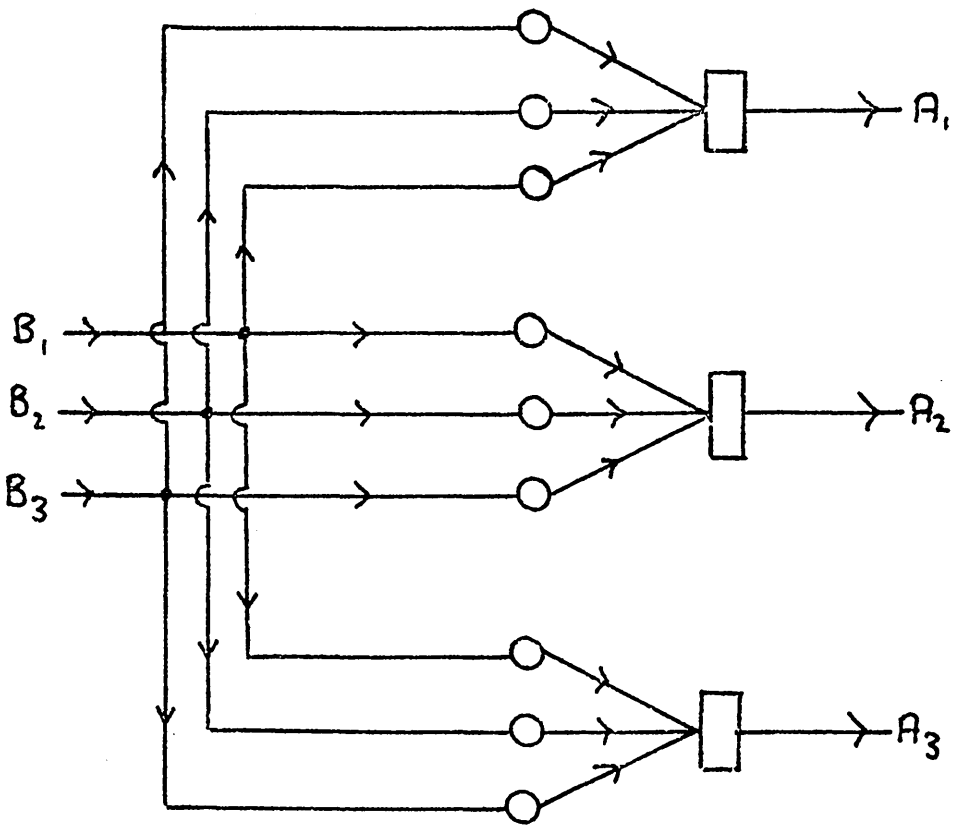

FIGURE 18

A 3x3 Associative Net as a battery of 3 Adalines

○ represents an amplifier.

▯ represents a threshold device.

as a battery of linear Perceptrons. There is a small difference in that, unlike the case of the Perceptron, only one of a small number of values may be assigned to each switch (or node) of a Net. In some Nets this number is as few as two. The important distinction is in the way values are assigned to these memory elements. Here we refer to our definitions of Classical and Operant Conditioning (section 1.2, page 2). A memory Net, by virtue of being an associative memory, stores information by the mechanism of Classical Conditioning. On the other hand, much of the interest in Perceptron work has focussed on the way in which the Perceptron can be taught to discriminate between sets of patterns by the reinforcement of correct responses and making incorrect responses less likely to occur in the future. Consequently this is an Operant Conditioning system.

8.3 Weight Adjustment Procedures employed in the memory Nets

As a result of considering memory Nets and Perceptrons we have, as is often done in Perceptron studies, looked at the procedures we have used for assigning values to memory elements of the Nets

There are many different procedures which can be used. The one used in the linear W.S.A.N. is infallible in the trivial case of the R A-messages stored being represented by an / . .

an orthogonal set of vectors. Consider the ith output line of a Net and let the values (weights) assigned to its N_B memory elements be contained in the vector $\underline{W}^{(i)}$. There are R sets of message pairs $(A_{ir},\underline{B}^{(r)})$ (r=1,2,..R) stored and, for every $\underline{B}^{(r)}$ input in the retrieval process, the system is required to output a pulse of strength proportional to the appropriate A_{ir}. $((A_{ir},\underline{B}^{(r)})$ is written for $(A_{ir},B_{1r}),(A_{ir},B_{2r}),.....(A_{ir},B_{N_Br}))$. Let the values of the N_B weights contained in $\underline{W}^{(i)}$ be constructed as follows.

$$\underline{W}^{(i)} = \sum_{r=1}^{R} A_{ir}\,\underline{B}^{(r)} \quad \ldots\ldots 8.1$$

Then, on input of message $\underline{B}^{(n)}$, the output to be fed into the threshold detector placed on the line A_i is

$$\underline{W}^{(i)}.\underline{A}^{(n)}$$

$$= \sum_{r=1}^{R} A_{ir}\,\underline{B}^{(r)}.\underline{B}^{(n)}$$

$$= A_{in} \sum_{j=1}^{N_B} B_{jn}^2$$

$$\propto A_{in}$$

if the vectors $B^{(r)}$ are mutually orthogonal. Thus the required classification is always attained. When the messages are vectors with components of value +1 or -1 and / . .

and the threshold θ is set at 0, reference to equation 7.19 (page 149) shows that this method is the one used in the W.S.A.N. The method for switching on nodes in the Associative Net, which deals with messages with components of value 0 or 1, is in general different from that described by equation 8.1 but, as it happens, since R messages represented by a set of mutually orthogonal vectors never activate any input line more than once in storage, equation 8.1 does describe the storage process in that case.

In general, however, the memories do not deal with orthogonal sets of vectors. A way of seeing how good the weight adjustment procedures we have employed are is to investigate how they compare with the best that can be achieved. In this connection we are able to say something about the Symmetrical Associative Net, which is an array of binary switches storing vectors with components +1 or -1. Once more we consider one of its N_A output lines.

There are 2^{N_B} B-messages which may be stored. Each may be put into one of two categories so that there are $2^{2^{N_B}}$ different ways of classifying them. There are, however, 2^{N_B} different weight vectors associated with this output line, each of which classifies the ensemble of 2^{N_B} vectors differently. This can be shown by proving / . .

proving that for any two weight vectors $\underline{U}$ and $\underline{V}$ there is at least one of the 2^{N_B} possible messages which they classify differently.

Suppose the two weight vectors differ in m of their N_B components which are ordered so that these m are identified by the indices 1 to m. We will try and find a vector $\underline{B}$ which is assigned to the +1 category by $\underline{U}$ and to the -1 category by $\underline{V}$. With the threshold θ set at 0 we thus require that

$$\underline{B}.\underline{U} > 0 \text{ and } \underline{B}.\underline{V} < 0 \quad \ldots\ldots 8.2$$

(N_A is an odd number so that neither of these scalar products ever equals 0)

Now $U_i = -V_i$ for i=1,2,..m

and $U_i = V_i$ for i=m+1,m+2, ... N_B.

Hence $$\underline{B}.\underline{U} = \sum_{i=1}^{m} B_i U_i + \sum_{i=m+1}^{N_B} B_i U_i$$

and $$\underline{B}.\underline{V} = -\sum_{i=1}^{m} B_i U_i + \sum_{i=m+1}^{N_B} B_i U_i$$

Writing $$S_m = \sum_{i=1}^{m} B_i U_i$$

and $$S'_{N_B} = \sum_{i=m+1}^{N_B} B_i U_i$$

then from 8.2, the conditions to be satisfied are

S_m / . .

$$S_m + S'_{N_B} > 0,$$

$$-S_m + S'_{N_B} < 0$$

i.e. $S_m > S'_{N_B}$, $S_m > -S'_{N_B}$

or $$S_m > S'_{N_B} > -S_m \quad \ldots\ldots\ldots\ldots 8.3$$

The coefficients of $\underline{B}$ can be chosen so that S_m has any integral value between +m and -m and, independent of this, S'_{N_B} has any integral value between N_B-m and $-(N_B-m)$, Let them be chosen so that $S_m=m$. Then it is clear, except in one case, that S_{N_B} may be chosen to have a non-negative value less than m, thus satisfying 8.3. The case which is not clear is when m=1, for then S'_{N_B} must have the value 0. Here, let B_1 be chosen so that $S_m=1$. Remembering that N_A is an odd number, S'_{N_B} is now a sum of an even number of terms and so the N_B-1 remaining components of $\underline{B}$ can be chosen so that $S'_{N_B}=0$.

We conclude that for any pair of binary weight vectors $\underline{U}$ and $\underline{V}$ there is always one vector out of the ensemble of the 2^{N_B} available which $\underline{U}$ and $\underline{V}$ classify differently. Thus each of the 2^{N_B} possible arrangements of the weights of this particular output line is identified with a different classification of this ensemble.

We will now calculate the probability that R arbitrarily chosen vectors can be classified correctly by some arrangement of the N_B weights.

There / . .

There are $2^{2^{N_B}}$ possible classifications of the ensemble of 2^{N_B} vectors, of which 2^{N_B} only are realisable in this situation. Assuming that those which are realisable are distributed randomly over the $2^{2^{N_B}}$ possible, the probability that a particular classification is realisable is

$$\frac{2^{N_B}}{2^{2^{N_B}}}.$$

If the memory is required to store R vectors, then there are $2^{2^{N_B}-R}$ classifications of the whole ensemble in which the R vectors are classified in a particular way. The probability that at least one of these is one of the 2^{N_B} realisable is thus

$$P = 1 - \left(1 - \frac{2^{N_B}}{2^{2^{N_B}}}\right)^{2^{2^{N_B}-R}}$$

$$= 1 - \exp(-2^{N_B-R}) \qquad \text{for } 2^{N_B} \gg R.$$

If a message is as likely to be put into category +1 as category -1 then R bits of information have been stored in N_B binary registers, and the information storage density is

$$D = R/N_B.$$

Writing $D=1-\delta$ then the probability of incorrect classification is

$$1-P = \exp(-2^{\delta N_B}).$$

For / . .

For any storage density less than 1 the probability of incorrect retrieval of the R bits of information may be made to be as small as possible by judicious choice of N_B, and as N_B approaches infinity the efficiency of retrieving information can be made to approach 1.

We see that the Symmetrical Associative Net can be made to work efficiently, with the only bar being that imposed by information theoretical considerations, provided the appropriate weight adjusting procedure is known. It is difficult to make a similar statement for the Associative Net. Although each of the 2^{N_B} weight vectors classify differently the ensemble of possible vectors whose components now take the value 1 or 0, the 2^{N_B} classifications which are realised can not be regarded as being distributed at random amongst the ensemble of possible classifications of messages to be stored. Certainly the Associative Net can be made to work with an information efficiency of 69% of the theoretical maximum. But there may well be other weight adjustment procedures which improves on this.

8.4 The memory Nets and Information Retrieval Systems

As far as is known, the properties of the memory Nets have not been explored with reference to the design of computer stores, despite the interest in content addressable and associative memories. The reason may be that since / . .

since this sort of distributed memory may give rise to errors in retrieval, even though the elements out of which it is made function correctly, it is not reliable enough for use as a hardware device in digital computers.

There is a parallel between an Associative Net and a technique of retrieval of information stored on punched cards (Mooers, 1954). Here we take one output line of an Associative Net to represent a card, the pattern of active nodes on that line being equivalent to the pattern of holes on the card. A document represented by a card is characterised by k descriptors, each of which is represented by punching holes at a particular N of the possible F sites of a card. The sites to be punched are selected pseudo-randomly. We use the system by asking for those cards which are characterised by a particular set of k descriptors. This is done by activating in turn the k sets of N input lines identified with the k descriptors. With a threshold on the output lines set at the number of active input lines, the cards which are obtained are those identified with the output lines which fire on all k occasions.

Mooers calculated the information retrieved from one card for each of the k descriptors, by looking at the card's F sites to see how probable it was that the descriptor had contributed to the marking of the card.

He / . .

He maximised the information gained (R_t) as a function of k, and by approximating $\frac{\partial R_t}{\partial k}$ for the case when the ratio $\frac{N}{F}$ could be regarded as small compared with 1, showed that R_t had a maximum value of ln2 bits per site when the density of holes on each card was 0.5. This is in line with our calculations (section 4.4) although, instead of looking at the states of individual nodes, we calculated the information gained from a large number of output lines by observing the final output from them. However, Mooers' system was not intended to be an Associative Net - for example since one card can be identified with one output line of the Net, the punched card system is not a distributed store. As far as we know, he has not considered the performance of his system when no approximations are made.

8.5 Gabor's model of associative memory

Stimulated by Longuet-Higgins' invention of the Holophone (Longuet-Higgins, 1968a), Gabor (1968a, 1968b) proposed two alternative transformations by which the holographic process could be imitated. He then went on to discuss (Gabor 1969), in a paper which predates our own on the Correlograph (Willshaw et al., 1969), a linear optical associative memory which is similar to the Linear Correlograph. These two models are not identical. Gabor's model deals with temporal signals and the information stored is not retrieved with maximum fidelity until / . .

until all of the cue has been fed into the system. Here plates A, B, C and S of the Linear Correlograph (shown in Figure 5, page 55) are replaced by moving film strips and the positions A, B and C occupy in Figure 5 are interchanged. Since the information stored is represented by sequences of 1's and -1's (as in the Holophone), there are two independent light tracks. Gabor considered the retrieval of a message containing 1's and -1's chosen with equal probability, by using a fragment of it as a cue. He showed how it is possible to use his device to recognise short fragments of coded sequences, but did not look at the model's performance in storing more than one message, nor did he calculate its information efficiency.

8.6 Conclusion

We have discussed the relationship between three systems and some of the distributed memory models we have considered. By virtue of our models' simple form there are, no doubt, other devices in existence which, like the Perceptron and the punched card system, are similar in structure to our own, although we doubt whether they would be also similar in function. There is, however, a close resemblance between our Correlographic models and Gabor's linear memory model. This is not surprising since, as we mentioned in Chapter 4, his observations / . .

observations on the possibilities of imitating the holographic process led us directly to construct the non-linear Correlograph.

CHAPTER 9

Biological analogues of the distributed memory models

9.1 Introduction

The main object of this work has been to construct efficient models of distributed memory and to enumerate their properties, with the hope of raising questions of neurophysiological interest. The models we have discussed have been presented in the temporal order in which they were constructed and an attempt has been made to show how they have developed one from another. We will now make a few remarks about their possible realisation in neurophysiological terms. The discussion will be brief as we feel that detailed investigation of neurophysiological implications should be left to those more competent in this respect. We will, however, pause to consider a relevant theory of the cerebellar cortex due to Marr which, as far as we can see, is founded on a large quantity of experimental data and requires few assumptions to be made. We shall show that in structure and in function Marr's model for the cerebellum is closely related to the Associative Net.

9.2 The Holophone

The known non-linearity of some neural responses and the fact that it has not been shown that stable frequency sensitive cells exist, indicates (although

there / . .

there is no conclusive evidence) that the Holophone may not be suitable for realisation in neural tissue. These remarks apply in equal force to other holographic theories of memory, including those (for example van Heerden 1963b, Pribram 1966, 1969; Westlake 1968) in which direct parallels between holograms and ensembles of interacting nerve cells have been drawn.

9.3 The Associative Net

We have tentatively suggested in Chapter 4 (section 4, page 71) that the Associative Net may be regarded as an axo-dendritic neural system. The N_B input lines in Figure 9 (page 68) are axons, the N_A output lines are dendrites and there is a modifiable synapse at each of the $N_A N_B$ intersections. The value of 10^6 that we have assigned to N_A and N_B in some of our calculations was thought to be in line with this neural model.

9.4 The Symmetrical Associative Net, the W.S.A.N. and the Correlograph

We cannot regard the Symmetrical Associative Net or the W.S.A.N. as a similar network of synaptic junctions, which are modified according to the numbers of excitatory signals (the positive pulses mentioned in Chapter 7) and inhibitory signals (negative pulses) influencing them, for this would involve dispensing with the concept of the on-off nature of the nerve impulse. We can conceive of / . .

of relatively simple axo-dendritic nets in which these difficulties are overcome and, furthermore, we can suggest a neural model of an Associative Net with tied switches to represent the logically complex Correlograph. However, such an enterprise which has no experimental backing, has little value. What we will do is to substantiate our neurophysiological claims for the Associative Net by relating it to a theory of the cerebellar cortex.

9.5 The cerebellar cortex and the Associative Net

Marr (Marr 1969) took up the suggestion of Brindley (Brindley 1964) that the cerebellum learns to perform motor skills so that subsequently it will function satisfactorily if given incomplete input information.

The structure of the cerebellum is known in considerable detail (Ramón y Cajal 1911, Eccles, Ito and Szentagothai 1967). It has two kinds of input, that via the mossy fibres and that via the climbing fibres, and one sort of output, from the inhibitory Purkinje cells which are related one-to-one to the climbing fibres. Each climbing fibre is connected polysynaptically to a Purkinje cell and is capable of firing it. The mossy fibres are mapped many to many onto the granular cells whose axons, the parallel fibres, pass through the dendritic trees of the Purkinje cells. There are also three other types of cell - the Golgi cell, the Stellate cell / . .

cell and the Basket cell. In effect, in Marr's model, each Purkinje cell, which is claimed to provide the instruction for an elemental movement of the organism, functions as one output line of an Associative Net. The input lines are the parallel fibres, each one making no more than one synaptic contact with any Purkinje cell dendrite tree, since these trees lie in planes perpendicular to the direction of the parallel fibres. Each input line is only connected to a selection of output lines. (Figure 19).

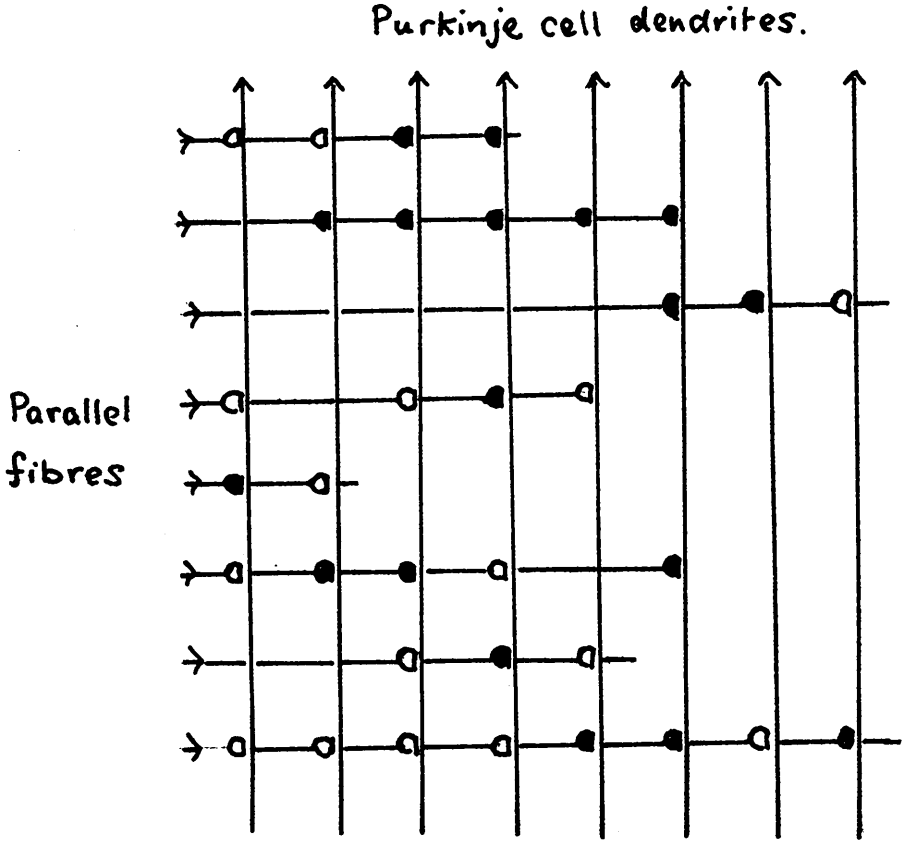

FIGURE 19
The storage system of the cerebellum as an Associative Net

He suggested that the mossy fibre input transmits the 'context' associated with the firing of a Purkinje cell by its climbing fibre. The synapses between the Purkinje cell dendrites and the parallel fibres activated by the mossy fibre input are facilitated so that on subsequent occasions, input along the mossy fibres alone is sufficient to excite the appropriate Purkinje cell. More than one context may be associated with the firing of the same Purkinje cell (just as in the Associative Net more than one input message will excite the same output line). Each Purkinje cell may be influenced by about 200,000 parallel fibres passing through its dendritic tree, which are activated by about 7,000 mossy fibres. The number of mossy fibres active at one time is assumed to vary between 20 and 200, and the number of parallel fibres consequently active (M) is set by the inhibitory Golgi cells which control the granular cell thresholds. In retrieval, the Purkinje cell thresholds are set by the Stellate cells which sample the parallel fibre activity. As a result, these thresholds are not able to be set at the number of active parallel fibres M crossing a dendrite tree, but at a value pM, where p is a number close to 1. In determining how this system may function optimally, it is required, unlike in the case of the Associative Net, that a lower limit be set on the value of M. This is to ensure that all the active / . .

active mossy fibres which form the input to the granular cells are represented in the granular cell input along the parallel fibres. Secondly, M must be large enough for a reliable sample of the parallel fibre activity to be taken by the Stellate cells. Marr observed that, provided these two conditions were met, M should be as small as possible in order that the number of contexts one Purkinje cell would be able to learn, without there being a large probability of error in its response, should be as large as possible. He calculated that if M is set at approximately 500 (but no less than 500), with 70% of the synapses of a Purkinje cell dendrite tree modified, the probability of a Purkinje cell responding when it should not, or remaining inactive when it should fire, is in each case no greater than 0.01. Under these conditions about 200 different contexts could be learned by each Purkinje cell. Apart from the complexities introduced because the number of active parallel fibres may vary subject to a lower limit and the Purkinje cell thresholds are set by sampling techniques, the crucial point of difference between Marr's analysis and that for the Associative Net is that the output from each Purkinje cell is supposed to initiate an elementary movement of the organism, independently of what other Purkinje cells are doing. As a consequence he did not consider the collective behaviour of a set of Purkinje / . .

Purkinje cells and so did not reach any conclusion about the cerebellum's information efficiency. We have shown that an Associative Net with constant thresholds on the output performs efficiently; Marr's model should be able to do likewise, although the efficiency may be adversely affected by the relatively large number of parallel fibres active in the stimulation of a particular Purkinje cell.

9.6 Conclusion

Although Marr's theory rests on the unverified hypothesis that the parallel fibre/Purkinje dendrite synapses are modifiable, he does relate closely his theoretical concepts to the large amount of experimental detail concerning the cerebellum.

For that reason, and since we have shown that the underlying logical principles of the Associative Net and the neural representation of it which we have suggested do have close ties with his cerebellal model,we have some confidence that at least one of the distributed models of memory that we have proposed may have neurophysiological significance.

CHAPTER 10

Conclusions

The mathematical relationships between the distributed models of memory that we have investigated will be summarised by relating our models to the linear/non-linear dichotomy we discussed in section 1.6.

Equations 1.1 and 1.2, which relate the output α to the input β of a class of memory models, are

$$\alpha = M^{-1}DM\beta \quad \ldots\ldots\ldots\ldots\ldots (1.1)\text{(linear)}$$

$$\alpha = [[M^{-1}DM]\beta] \quad \ldots\ldots\ldots (1.2)\text{(non-linear)}$$

We take α to be an N_A-vector, β an N_B-vector. We choose M to be the square discrete Fourier Transform matrix with components $M_{jk} = \frac{1}{N}\exp(\frac{2\pi i}{N}jk)$, where N equals N_A or N_B as appropriate. M^{-1} is the inverse of M. The $N_A x N_B$ matrix D initially has components $D_{pq} = \delta_{pq}$, so that before storage the output is either a simple linear (equation 1.1) or a non-linear (equation 1.2) function of the input.

In all cases we will suppose that R pairs of messages $(A^{(r)},B^{(r)})$ (r=1,2,...R), represented by the R columns of the matrices A and B, are to be stored in the model under consideration. The components of A and B are only allowed to take the values 1 or 0.

The Correlograph

The R message pairs $(A^{(r)},B^{(r)})$ are stored by choosing / . .

choosing D to have components

$$D_{pq} = \delta_{pq} \sum_{r=1}^{R} \phi^{A}_{pr} \phi^{B*}_{qr}, \quad \ldots\ldots\ldots 10.1$$

where ϕ^{A}=MA and ϕ^{B}=MB.

We have shown in section 4.5 that, for this situation, equation 1.2 is applicable to the non-linear Correlograph.

The Holophone

We now choose A and B to be identical and construct a new matrix AA, to be used in place of A (and B). The components of A are mapped one-to-one onto the components of AA - namely $AA_{ij} = 2A_{ij}-1$. AA_{ij} can thus have the value +1 or -1. The matrix D is constructed as for the Correlograph, except that in equation 10.1 AA replaces A and B. The stored message pairs are thus $(AA^{(r)},AA^{(r)})$ and a fragment $AA'^{(n)}$ is used to retrieve the rest of $AA^{(n)}$ from the memory. Equation 1.1 becomes

$$\alpha = M^{-1}DMAA'^{(n)}$$

This is identical with equation 3.14 (page 34) and thus describes the functioning of the linear Holophone.

The Off-diagonal Holophone

D is now allowed to have non-zero components in off-diagonal positions, namely

$$D_{pq} = \sum_{r=1}^{R} \phi^{A}_{pr} \phi^{B*}_{qr} \quad \ldots\ldots\ldots 10.2$$

If, as for the Holophone, we store the message pairs $(AA^{(r)},BB^{(r)})$ / . .

$(AA^{(r)},BB^{(r)})$, whose components are linearly related to those of $(A^{(r)},A^{(r)})$ then, as we demonstrated in section 3.7 (page 46), equation 1.1 now describes the behaviour of the linear Off-diagonal Holophone.

The Symmetrical Associative Net

In the case of the Symmetrical Associative Net, which stores message pairs of the form $(AA^{(r)},BB^{(r)})$ where $BB_{ij}=2B_{ij}-1$ and A and B are chosen independently, then with D as chosen for the Off-diagonal Holophone we have shown in Chapter 7, (page 135) that this is a non-linear model, described by equation 1.2.

The W.S.A.N.

This is identical to the linear Off-diagonal Holophone except that a non-linear operation is performed in the retrieval process. Although it does not fit into our linear/non-linear dichotomy, we shall call it a linear model after the manner in which its store is set up. It is best described by the equation

$$\alpha = [\, M^{-1}DM\beta \,]$$

The Associative Net

Finally if, as for the Correlograph, A and B are chosen independently and message pairs $(A^{(r)},B^{(r)})$ are stored directly then, as we saw in section 4.5 (page 73A), the mathematics of the non-linear Associative Net are summarised in equation 1.2.

We / . .

We have considered in some detail five of the family of memory models which are described by equations 1.1 and 1.2. All are distributed associative memories which are distinguished one from another by the form of their memory matrix D, whether the information is handled directly or firstly undergoes a simple encoding operation and whether they are linear or non-linear systems (the W.S.A.N., which we have called a linear model is in fact not strictly described by equation 1.1) The non-linear models considered possess the advantage of having relatively simple logical structure. This enabled us to make general statements about their behaviour and led us to investigate in detail the properties of two of them - the Correlograph and the Associative Net. It was shown by analytical and computer simulation methods that, if they are allowed to deal with highly redundant messages, then they can be made to work efficiently in storing and retrieving information when compared with the best that can be expected from information theoretical considerations. Furthermore, it was shown how they can be modified to perform a wide range of relevant tasks efficiently. In particular, we considered the biologically important questions of how the memories could function in the face of damage to the store locations or to the address messages used in recall. We looked at the models acting as content-addressable devices in which an arbitrary

part / . .

part of a stored message was used to retrieve the rest of it from store and we showed how the Associative Net could learn a sequence of instructions with the aid of a feedback mechansim. We speculated on a way by which the Associative Net could be used indefinitely and analysed the behaviour of these models acting as recognition devices. With regard to the Symmetrical Associative Net which was the other non-linear model that we looked at, we were able to show that it could perform relatively efficiently, whether damaged or undamaged, when it is dealing with messages containing no redundancy. In all three cases we have been careful to point out that there is no guarantee, even if the system is working at its maximum efficiency, that a message will be retrieved perfectly. This is due to the nature of a distributed store, in which each location may not be identified with the storage of just one message pair. In the three models we have considered the fractional information loss at maximum efficiency is small, and can be reduced to whatever level is required by loading the system more lightly, thereby decreasing the information efficiency.

In order to limit our problem we have had to make many assumptions and many topics which border on this work have had to be left aside. We have employed information theory to tell us how good our memory models are. We do, however, realise that efficient usage of storage / . .

storage space may not be the only desirable property of a memory. For example, it may be important to minimise the number of elementary operations that a memory uses in performing a particular task.

Since in general we have not had a concrete physical model in mind, we have concentrated on finding the optimum conditions under which our memories may perform particular tasks, rather than demonstrating in detail how these conditions are achieved. For example, we have assumed that the messages to be processed may be represented as binary vectors of the right length, each containing the correct number of 1's and 0's, without explaining how we envisage these coding operations are to be carried out.

Our study of memory has, however, not been completely abstract. From time to time we have been guided, in particular in our work on the Associative Net, by keeping an eye on neurophysiological implications. In fact, as we saw in Chapter 9, it is the translation of this model into neural terms which gives us hope that at least some of our models of distributed associative memory have biological relevance.

REFERENCES

BEURLE, R.L. (1956). Properties of a mass of cells capable of regenerating pulses. Phil.Trans.R.Soc. Lond. B 240, 55.

BLOCK, H.D. (1962). The Perceptron: a model for brain functioning, I. Rev.Mod.Phys. 34, 123.

BRINDLEY, G.S. (1964). The use made by the cerebellum of the information that it receives from sense organs. Int.Brain.Res.Org.Bull. 3, 80.

COLLIER, R.J. (1966). Some current views on holography. I.E.E.E. Spectrum 3, 67.

CRAGG, E.C. & TEMPERLEY, H.N.V. (1954). The organisation of neurones; a co-operative analogy. Electroenceph. clin.Neurophysiol. 6, 85.

ECCLES, J.C., ITO, M & SZENTAGOTHAI, J. (1967). The Cerebellum as a Neuronal Machine. Berlin: Springer-Verlag.

FARLEY, B.G. & CLARK, W.A. (1954). Simulation of self organising systems by digital computer. Trans.IRE. Information Theory.

GABOR, D. (1948). A new microscopic principle. Nature. Lond. 161, 777.

GABOR, D. (1949). Microscopy by reconstructed wavefronts. Proc.R.Soc.Lond. A 197, 187.

GABOR, D. (1951). Microscopy by reconstructed wavefronts. II. Proc.phys.Soc. 64, 449.

GABOR, D. (1968a). Holographic model of temporal recall. Nature.Lond. 217, 584.

GABOR, D. (1968b). Improved holographic model of temporal recall. Nature, Lond. 217, 1288.

GABOR, D. (1969). Associative holographic memories. IBM. Jl Res.Dev. 13, 156.

GOMULICKI, B.R. (1953). The development and present status of the trace theory of memory. Brit.J.Psychol. monograph supplements 29.

HEBB, D.O. (1949). The Organisation of Behaviour. New York: Wiley.

VAN HEERDEN, P.J. (1963a). A new optical method of storing and retrieving information. Appl.Optics. 2, 387.

VAN HEERDEN, P.J. (1963b). Theory of optical information storage in solids. Appl.Optics. 2, 393.

VAN HEERDEN, P.J. (1970). Models for the brain. Nature, Lond. 225, 177.

HOPGOOD, F.R.A. (1969). Compiling Techniques. London: Macdonald.

KOFFKA, K. (1935). Principles of Gestalt Psychology. New York: Harcourt, Brace & Co.

KÖHLER, W. (1940). Dynamics in Psychology. New York; Liveright Publ.Corp.

LASHLEY, K.S. (1942). The problem of cerebral organisation in vision. In Visual Mechanisms. H.Klüver ed. Lancaster Pa: Jacques Cattell Press.

LEITH, E.N. & UPATNIEKS, J. (1962). Reconstructed wavefronts and communication theory. J.opt.Soc.Am. 52, 1123.

LONGUET-HIGGINS, H.C. (1968a). Holographic model of temporal recall. Nature, Lond. 217, 104.

LONGUET-HIGGINS, H.C. (1968b). The non-local storage of temporal information. Proc.R.Soc.Lond. B 171, 327.

LONGUET-HIGGINS, H.C., WILLSHAW, D.J. & BUNEMAN, O.P.(1970) Theories of associative recall. Qu.Rev.Biophys. 3, 223.

MARR, D. (1969). A theory of cerebellar cortex. J.Physiol. Lond. 202, 437.

MILNER, P.M. (1957). The cell assembly: Mark II. Psychol. Rev. 64, 242.

MINSKY, M & PAPERT, S. (1969). Perceptrons. Boston: M.I.T. Press.

MOOERS, C.N. (1954). Choice and coding in information retrieval systems. In IRE.Trans. 1954 Symposium on Information Theory.

PAVLOV, I.P. (1927). Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. London: Oxford University Press.

PLATO. The Theaetetus. translated by S.W. Dyde (1899). Glasgow: Maclehose.

PRIBRAM, K.H. (1966). Some dimensions of remembering: steps towards a neuropsychological model of memory. In Macromolecules and Behaviour (ed. J.Gaito). New York: Appleton-Century-Crofts.

PRIBRAM, K.H. (1969). The neurophysiology of remembering. Scient.Am. 220, 73.

RAMÓN Y CAJAL (1911). Histologie du Système Nerveux de l'Homme et des Vertébrés. Paris: A.Maloine.

RASHEVSKY, N. (1938). Mathematical Biophysics. Chicago: University of Chicago Press.

ROCHESTER, N. HOLLAND, J.H., HAIBT, L.H. & DUDA, W.L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. IRE Trans. Information Theory. IT-2, 80.

ROSENBLATT, F. (1958). The perceptron. A probabilistic model for information storage and organisation in the brain. Psychol.Rev. 65, 386.

ROSENBLATT, F. (1962). Principles of Neurodynamics. Washington D.C.: Spartan Books.

ROY, A.E. (1960). On a method of storing information. Bull.math.Biophys. 22, 139.

ROY, A.E. (1962). On a method of storing information II. A further study of model properties. Bull.math. Biophys. 24, 39.

STROKE, G.W. (1966). An Introduction to Coherent Optics and Holography. New York: Academic Press.

THORNDIKE, E.L. (1949). Selected Writings from a Connectionist's Psychology. New York: Appleton-Century-Crofts.

WESTLAKE, P.R. (1968). Towards a theory of brain functioning: a detailed investigation of the possibilities of neural holographic processes. Ph.D. Dissertation. U.C.L.A.

WIDROW, B. (1964). Pattern recognition and adaptive control. Trans.A.I.E.E.E. Appl.Indust. 83, 269.

WILLSHAW, D.J., BUNEMAN, O.P. & LONGUET-HIGGINS, H.C. (1969). Non-holographic associative memory. Nature, Lond. 222, 960.

WILLSHAW, D.J., BUNEMAN, O.P. & LONGUET-HIGGINS, H.C. (1970). Models for the brain. Nature,Lond. 225, 178.

WILLSHAW, D.J. & LONGUET-HIGGINS, H.C. (1969). The holophone - recent developments. In Machine Intelligence 4. (ed. D.Michie). Edinburgh University Press.

WILLSHAW, D.J. & LONGUET-HIGGINS, H.C. (1970). Associative memory models. In Machine Intelligence 5. (ed.D.Michie) Edinburgh University Press.

YOUNG, J.Z. (1966). The Memory System of the Brain. Oxford University Press.

YOUNG, J.Z. (1970). What can we know about memory? Brit. Med.J. 1, 647.

PUBLISHED PAPERS

WILLSHAW, D.J. & LONGUET-HIGGINS, H.C. (1969). The holophone - recent developments. In Machine Intelligence 4. (ed. D.Michie). Edinburgh University Press.

WILLSHAW, D.J., BUNEMAN, O.P. & LONGUET-HIGGINS, H.C. (1969). Non-holographic associative memory. Nature, Lond. 222, 960.

WILLSHAW, D.J. & LONGUET-HIGGINS, H.C. (1970). Associative memory models. In Machine Intelligence 5. (ed. D.Michie) Edinburgh University Press.

WILLSHAW, D.J., BUNEMAN, O.P. & LONGUET-HIGGINS, H.C. (1970). Models for the brain. Nature, Lond. 225, 178.

LONGUET-HIGGINS, H.C., WILLSHAW, D.J. & BUNEMAN, O.P. (1970). Theories of associative recall. Qu.Rev. Biophys. 3, 223.

19

The Holophone - Recent Developments

D. J. Willshaw
and
H. C. Longuet-Higgins
Department of Machine Intelligence and Perception
University of Edinburgh

In this paper we review some of the properties of the holophone (Longuet-Higgins 1968a and b), which is a device analogous to the holograph (Collier 1966) but working in time rather than in space. It was invented to illustrate the principle of non-local information storage as applied to temporal signals, a principle which may very well be used in the human brain. But whether or not this is so, it seems worth while to explore the behaviour of the holophone in some detail, since the device might find application in man-made memory systems.

Before embarking on mathematical details, it may be helpful to indicate some of the properties of the holophone, viewed as a black box with one input channel and one output channel. Three properties are of special interest:

1. It can be used to record any input signal which lies in a certain frequency range and does not exceed a certain length. If part of a recorded signal is then put into the holophone, the continuation of the signal emerges, in real time. In this paper we investigate the amount of noise associated with the playback.
2. Several signals can be recorded on the same holophone. If the signals are random, an input cue from one of the signals will evoke the continuation of that same signal. The accompanying noise increases with the number of recorded signals.
3. The holophone can be used, like an optical filtering system, for detecting the occurrence of a given segment in the course of a long signal. What one does is to record on the holophone the segment of interest followed immediately by a strong pulse. The long signal is then played into the holophone; immediately after an occurrence of the recorded segment a pulse will emerge from the holophone. This property has not been discussed before, and we shall present the relevant theory.

In essence the holophone is a bank of narrow-pass filters, connected in parallel to an input channel, and also connected in parallel, through amplifiers of variable gain, to an output channel. The memory of the system resides in the gains of the various amplifiers. Figure 1 illustrates the layout of the system.

The recording of an input signal is carried out in two stages. The first stage, which corresponds to the formation of a latent image in photography, is to measure the power transmitted by each filter during the passage of the signal. This calls for a set of integrators, which are not shown in figure 1.

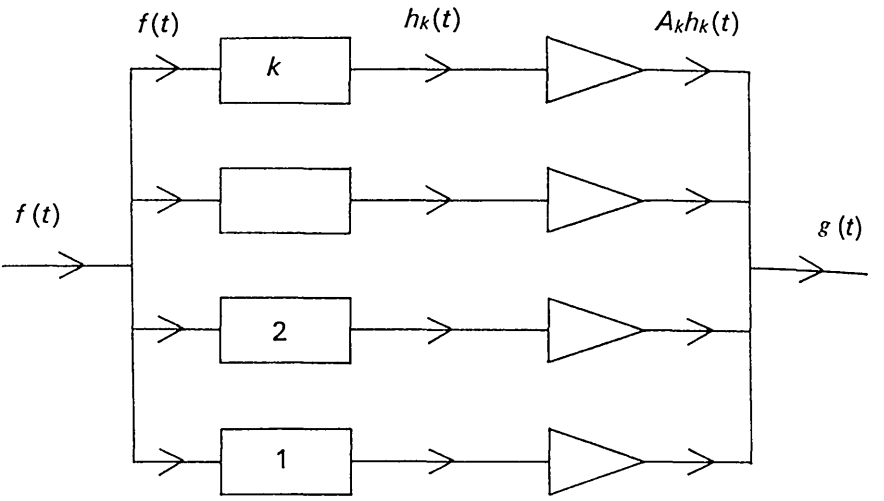

Figure 1

The second stage, corresponding to photographic development, is to turn up the gain of each amplifier by an amount proportional to the value stored in the corresponding integrator. The overall result is to change the response function of the holophone by an amount depending on the temporal auto-correlation of the recorded signal, and this is the secret of the device. But these cursory remarks are unlikely to carry conviction without further argument, so we now give a brief account of the underlying mathematical theory.

Let $f(t)$ be an input signal and let $h_k(t)$ be the output of the kth filter. Each filter must respond linearly

$$h_k(t) = \int_0^\infty R_k(\tau) f(t-\tau)\, d\tau,$$

and its response function must be of the form

$$R_k(\tau) = \frac{\mu}{\pi} e^{-\mu\tau} \cos k\mu\tau.$$

The quantity μ in this expression represents both the bandwidth of every filter and the spacing between the resonant frequencies of neighbouring filters, so that the given frequency range is fully covered. For a particular setting of the amplifiers the output signal $g(t)$ is given by

$$g(t) = \sum_k A_k h_k(t),$$

where A_k is the gain of the kth amplifier.

Suppose now that we wish to record a signal $f(t)$ which is over by the time $t=0$. We arrange for the integrators to measure the quantities

$$W_k(f) = \int_{-\infty}^{0} f(t) e^{2\mu t} h_k(t)\, dt,$$

which may be thought of as the amounts of work done by $f(t)$ upon the various filters, with greater weight attaching to the more recent events. (It can be shown that the W_k are essentially positive quantities, a point of importance for what follows.) Subsequently, at leisure, we increase each gain A_k by a proportional amount, namely

$$\Delta A_k = (2\pi\lambda/\mu) W_k.$$

This process has the effect of altering the response function of the holophone, defined by the equation

$$g(t) = \int_0^{\infty} M(\tau) f(t-\tau)\, d\tau.$$

Detailed analysis shows that when μ is small the change in M due to the recording of f is

$$\Delta M(\tau) = \lambda \int_{-\infty}^{0} f(t') e^{2\mu(t'-\tau)} f(t'-\tau)\, dt'.$$

This is a time-weighted autocorrelation integral of the recorded signal. If the duration of f is short compared to μ^{-1}, the exponential term in the integrand may be neglected; if it is much longer the earlier part of the signal will be forgotten. The quantity μ^{-1} therefore sets an effective upper limit on the length of signal that can be recorded.

Suppose that initially all the amplifier gains are zero, so that $M(\tau) \equiv 0$, and that a signal f is then recorded. After the recording the response function of the holophone will be given by the above expression for $\Delta M(\tau)$, and a new input signal f' will give rise to an output

$$\begin{aligned} g(t) &= \lambda \int_0^{\infty} \Delta M(\tau) f'(t-\tau)\, d\tau \\ &= \lambda \int_0^{\infty} \left[\int_{-\infty}^{0} f(t') e^{2\mu(t'-\tau)} f(t'-\tau)\, dt' \right] f'(t-\tau)\, d\tau. \end{aligned}$$

It might be supposed that this output is merely an indistinct 'echo' of f', and in general this will be the case. But if f is a sufficiently complicated signal, and if f' happens to be an excerpt from it, a different conclusion must be drawn. To see why, let us begin by writing $g(t)$ in the alternative form

$$g(t) = \lambda \int_{-\infty}^{0} f(t')C(t,t')\,\mathrm{d}t',$$

where $$C(t,t') = \int_{0}^{\infty} f(t'-\tau)\,\mathrm{e}^{2\mu(t'-\tau)} f'(t-\tau)\,\mathrm{d}\tau.$$

For times t after the end of the cue f', the integral C becomes a function only of $t-t'$:

$$C(t-t') = \int_{-\infty}^{\infty} f(t'-t+s)\,\mathrm{e}^{2\mu(t'-t+s)} f'(s)\,\mathrm{d}s.$$

It then represents (see figure 2) the degree of resemblance between the cue f' and the section of f that was played into the holophone $t-t'$ units of time ago. If f is sufficiently irregular, $C(t-t')$ will be small unless $t-t'$ equals some fixed time interval θ. We deduce that in these circumstances

$$g'(t) = \lambda \int_{-\infty}^{0} f(t')C(t-t')\,\mathrm{d}t' \propto f(t-\theta),$$

so that the output g' will approximate to a continuation of the recorded signal f, carrying on from the moment at which the cue comes to an end. This is our first important result, anticipated at the beginning of the paper.

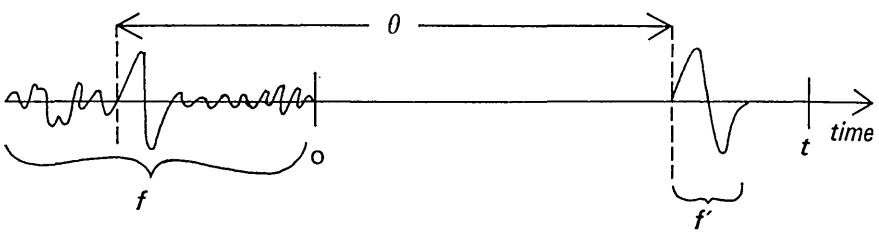

Figure 2

The recall of a whole recorded signal by presentation of an excerpt from it is analogous to the phenomenon of 'ghosts' in holography. Two objects are illuminated by the same coherent light source, and the scattered wavefronts are made to produce an interference pattern on a photographic plate. One of the objects is removed, and the other is illuminated as before and viewed through the interference pattern. A ghost of the absent object is seen beside the object which is actually there. In the temporal case the recorded signal f represents both objects, while the cue f' represents the object which was left

in position during the viewing process. There is only one non-trivial difference between the two cases: the cue f' can only evoke that part of the recorded signal which followed it, not the part which preceded it. It is possible, however, to evoke a time-reversed form of the earlier part of f by playing the cue in backwards! The interested reader may care to establish this curious property for himself.

Before turning to the question of noise in the playback we shall explain how the holophone can be used as a recognition device, since this promises to be one of its most useful applications. Suppose that we are faced with the problem of detecting the occurrence of a short segment f' in the course of a long signal f. (Both f and f' are assumed to be 'noise-like', having no marked periodicities.) What we do is to record on the holophone a signal consisting of f' followed immediately by a strong pulse at $t=0$. The response function of the holophone will then become

$$M(\tau) = \lambda \int_{-\infty}^{0} [\delta(t') + f'(t')] e^{2\mu(t'-\tau)} [\delta(t'-\tau) + f'(t'-\tau)] \, dt'.$$

Expansion yields four terms, of which the third vanishes because $f'(\tau)=0$ for positive τ and the fourth may be neglected if the pulse $\delta(t')$ was strong enough compared with the segment $f'(t')$. On this assumption

$$M(\tau) = \lambda[\delta(\tau) + e^{-2\mu\tau} f'(-\tau)]$$

so that $$g(t) = \lambda f(t) + \lambda \int_{0}^{\infty} e^{-2\mu\tau} f'(-\tau) f(t-\tau) \, d\tau.$$

In this expression for the output evoked by $f(t)$, the first term is uninteresting, being merely a playout of f itself. But the other term is a weighted correlation between f and f', and will make a sudden sharp contribution to the output whenever the recently received section of f is identical with the recorded segment f'. The prepared holophone therefore emits a sharp pulse immediately after any occurrence of the segment which it has been designed to detect. This property of the holophone is precisely analogous to the use of holography in the detection of special features such as printed words in an extended spatial pattern such as a page of a book.

An important question about the holophone is: how much noise will accompany the playback evoked by a cue taken from a recorded message? There is one special case which can be quickly disposed of, corresponding to the case of a collimated reference beam in holography. If the signal to be recorded consists of a strong pulse followed by a weaker signal of some sort, then after the recording has been completed the input of a pulse will evoke the weaker signal virtually free of noise. The reason is simple: in our earlier expression for $C(t-t')$, $f'(s)$ becomes $\delta(s-t_2)$, that is, a pulse at time t_2, and in

$f(t'-t+s)$ the only significant term is that arising from the recorded pulse, namely $\delta(t'-t+s-t_1)$, where $t_2-t_1=0$. Hence

$$C(t-t') = e^{2\mu(t'-t+t_2)}\delta(t'-t+0),$$

and our expression for $g(t)$ reduces to

$$g(t) = \lambda e^{2\mu t_1} f(t-0) \propto f(t-0).$$

But the more general case must be considered, and to this end we have reformulated the mathematics in discrete terms, assuming all signals to be short enough for the exponential decay factor to be neglected. For further convenience we have also imposed a cyclic boundary condition on the time dimension, and regarded each signal as a set of numbers associated with the vertices of a regular N-sided polygon. The recorded signal is then represented by a set of N numbers, each of which is assigned the value $+1$ or -1; the cue is taken to be a limited selection of L of these numbers at adjacent vertices, the other $N-L$ being assigned the value 0. No loss of generality is then suffered by writing the cue as

$$[f'_1, f'_2, \ldots, f'_N] = [f_1, f_2, \ldots, f_L, 0, \ldots, 0],$$

where the recorded signal is

$$[f_1, f_2, \ldots, f_N].$$

With these simplifications the following non-rigorous argument leads to a tentative expression for the signal-to-noise ratio of the 'playback' $[g_{L+1}, \ldots, g_N]$ evoked by the cue $[f'_1, \ldots, f_N]$. Defining C_m by the equation

$$C_m = \sum_n f'_n f_{n-m},$$

we may write g_i in the form

$$g_i = \sum_j f_j C_{i-j}.$$

The sum on the right-hand side includes N terms, one of which may be expected to be much larger than the others, namely that for which $i=j$. The value of C_0 is in fact just L, since each of the components of f' matches one of f. But every other term C_{i-j} is the sum of L elements each of which is $+1$ or -1 with equal probability (if the components are random). So taken together these terms have a variance equal to $(N-1)L$, while the square of the amplitude of the signal – the term in C_0 – is just L^2. The signal-to-noise ratio is therefore $L^2/(N-1)L$, which simplifies to L/N when N is large.

The above argument suggests that the signal-to-noise ratio should be approximately equal to the cue length divided by the length of the recorded signal (for long signals), but we thought it advisable to check the result by computer simulation. POP-2 was chosen as the program language. The following operations were carried out:

1. The N components of an input signal were generated with the aid of a pseudo-random number generator.
2. The first L of these were used to calculate the correlations C_m defined above.
3. The N–L components of the playback signal were then calculated according to the above equation for g_i.
4. Of these components approximately half arise from signal components equal to $+1$, and for this subset the signal-to-noise ratio was calculated from the formula

$$(S/N)_+ = (g_{av})^2/((g^2)_{av}-(g_{av})^2).$$

5. The same was done for the components arising from signal components equal to -1, and the two results were averaged to give an overall signal-to-noise ratio for the entire playback.

A sample set of results is shown below. A single 801-component signal had been recorded, and cues of varying length were used to recall it. The computed and theoretical values of the signal-to-noise ratio are tabulated against the number of components in the cue.

Length of cue (L)	S/N (computed)	S/N (theoretical)
150	0·214	0·187
200	0·252	0·250
250	0·301	0·313
300	0·396	0·375
350	0·412	0·466
400	0·513	0·500
450	0·553	0·572

The computed signal-to-noise ratios bear out the theoretical expression rather well in this case.

We also thought it advisable to test our theoretical estimate of the signal-to-noise ratio when several signals have been recorded on the holophone, and a cue from one of them is used to recall the rest of it. A straightforward extension of our earlier argument indicates that in this case the signal-to-noise ratio should equal the cue length divided by the combined length of all the recorded signals. To test this prediction we recorded ten signals, each of 151 components, and provided cues of varying length from arbitrarily selected signals. The results were as follows:

Length of cue (L)	S/N (computed)	S/N (theoretical)
31	0·0175	0·0206
46	0·0348	0·0303
61	0·0354	0·0404
76	0·0372	0·0503
91	0·0375	0·0604
106	0·0450	0·0701

Here the agreement is less good, so we checked our last three values by repeating the computation on a fresh set of signals, with the following results:

76	0·0450	0·0503
91	0·0436	0·0604
106	0·0500	0·0701

Presumably the discrepancy between the computed and the theoretical ratios is due to the non-independence of the various C_m, a feature which assumes greater importance for smaller values of N. Be that as it may, the computations show that the primitive theory (which assumes them independent) is at least roughly correct, and may be used as a basis for rough predictions about the behaviour of the holophone (and, for that matter, the holograph).

The above results show that the holophone will indeed function as a content-addressable memory, but that in this rôle it has rather distressing noise characteristics. Used as a recognition device, however, it should perform much more satisfactorily, and might even assume some technical importance. Let us briefly examine the theory of this process, using the same simplifications as were introduced earlier. Using $[f'_1,f'_2,\ldots,f'_L]$ to denote the signal which is to be recognized, and $[\ldots f_{-1},f_0,f_1,\ldots]$ to denote the input signal, we obtain the following simple expression for the detection signal:

$$\Delta g_i = \sum_{k=0}^{L-1} f_{i-k} f'_{L-k}.$$

If for some value of i the relationship

$$f_{i-k} = f'_{L-k}$$

holds for $k=0,\ldots,L-1$, then the ith component of the detection signal will be a spike of height L, and a spike of this height will signify with certainty the occurrence of f' in f.

A more interesting and realistic problem is that of detecting a slightly noisy version of f' in the longer signal f. The amount of noise in f can be specified by a parameter p which is the probability that the sign of a particular component of f' is wrongly quoted in the input signal f. For this noisy occurrence of f' to be detected, the threshold of the detection device must be lowered below the value L, but not too much or else it will emit false alarms. A sensible criterion for optimizing the threshold is to lower it until the increase in false-alarm probability equals the increase in probability of detection. On this criterion the optimum threshold T is found to have the value

$$T=L(1+2/\log_2 p),$$

when p is small and L is large.

Like the holograph, the holophone is a non-local, content-addressable storage and retrieval system. It further resembles the holograph in employing highly parallel logic and in being relatively invulnerable to damage of individual components. It was these characteristics which seemed to recommend it as a possible model of the human temporal memory – though in this context it must be viewed with all possible circumspection. The computations which we carried out to simulate its performance brought home to us the extreme difference in speed between the action of a holophone (which delivers its playback in real time) and the running of a computer program designed to simulate it. The difference is due, of course, to the fact that if the recorded signal is of length N, then N^3 separate acts of multiplication are needed to construct the output evoked by a cue. It is for this reason that the holophone, like the holograph, may be more useful as a hardware device than as a software subroutine – if it eventually finds a place in computing technology.

REFERENCES

Collier, R.J. (1966) Some Current Views on Holography. *IEEE Spectrum*, July 1966, 67–74.

Longuet-Higgins, H.C. (1968a) A Holographic Model of Temporal Recall. *Nature*, **217**, 104.

Longuet-Higgins, H.C. (1968b) The Non-Local Storage of Temporal Information. *Proc. Roy. Soc.* (in press).

(Reprinted from Nature, Vol. 222, No. 5197, pp. 960-962, June 7 1969)

Non-Holographic Associative Memory

by
D. J. WILLSHAW
O. P. BUNEMAN
H. C. LONGUET-HIGGINS
Department of Machine Intelligence and Perception, University of Edinburgh

The features of a hologram that commend it as a model of associative memory can be improved on by other devices.

THE remarkable properties of the hologram as an information store have led some people[1,2] to wonder whether the memory may not work on holographic principles. There are, however, certain difficulties with this hypothesis if the holographic analogy is pressed too far; how could the brain Fourier-analyse the incoming signals with sufficient accuracy, and how could it improve on the rather feeble signal-to-noise ratio[3] of the reconstructed signals? Our purpose here is to show that the most desirable features of holography are manifested by another type of associative memory, which might well have been evolved by the brain. A mathematical investigation of this non-holographic memory shows that in optimal conditions it has a capacity which is not far from the maximum permitted by information theory.

Our point of departure is Gabor's observation[4,5] that any physical system which can correlate (or for that matter convolve) pairs of patterns can mimic the performance of a Fourier holograph. Such a system, which could be set up in any school physics laboratory, is shown in Fig. 1. The apparatus is designed for making "correlograms" between pairs of pinhole patterns, and then using the correlogram and one of the patterns for reconstructing its partner. One of the pinhole patterns is mounted at A, and the other at B. The distance between them equals f, the focal length of the lens L. A viewing screen is placed at C, at a distance f from the lens, and a diffuse light source is mounted behind A. The pattern of bright dots appearing at C is the correlogram between the pattern at A and the pattern at B. Formally, $C = \bar{A} * B$, where the asterisk stands for convolution and $\bar{A}$ is the result of rotating the pattern A through half a turn round the optical axis. If A and B were interchanged, the pattern at C would be $\bar{B} * A = A * \bar{B} = \bar{C}$, so that the correlogram would be inverted. This is clear enough if B is a pinhole, and shows that the order of the patterns is important.

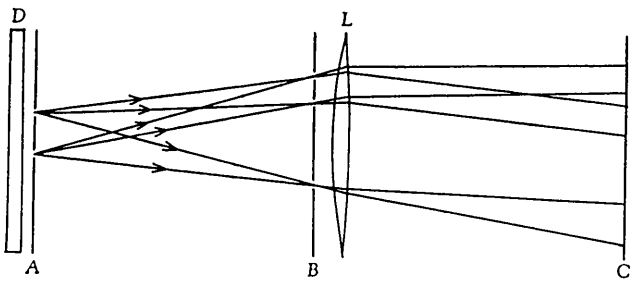

Fig. 1. Constructing a correlogram. D is a diffuse light source, L a lens and C the plane of the correlogram of A with B.

To recover pattern A from pattern B we convert the correlogram into a pattern of pinholes in a black card and place the light source behind it, so that the light shines through C and B on to a viewing screen at A (Fig. 2) A pattern of spots now appears on the viewing screen. All the spots of the original pattern A are present, but a number of spurious spots as well. If the pinholes were infinitesimal and there were no diffraction effects the reconstructed pattern would be $\bar{C} * B = A * \bar{B} * B$, just as in Fourier holography. If B were a random pattern, one could argue, $\bar{B} * B$ would approximate to a delta function at the origin, so that the reconstructed pattern would look like a slightly bespattered version of the original pattern A. How can we pick out the genuine spots from the others?

To solve this problem let us simplify the set-up by removing the lens (Fig. 3). Suppose, for example, that A has two holes and B has three. Then the pattern C will consist of six bright spots (barring coincidences). When these spots are converted into pinholes and illuminated from the right, a total of 18 ($= 6 \times 3$) rays will emerge from B and impinge on the screen at A. But we shall not see eighteen spots on this screen, because six of the rays will converge, in sets of three, on to the two points

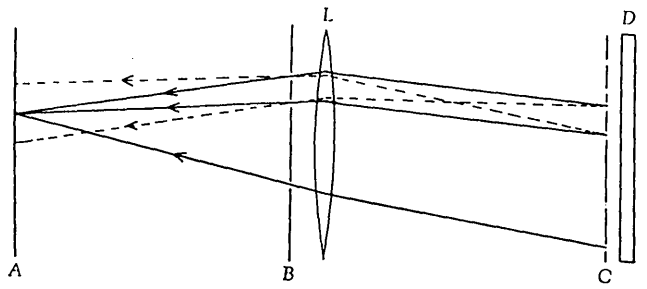

Fig. 2. Reconstructing a pattern. ———, Paths traversed in Fig. 1; - - -, paths not traversed in Fig. 1.

of the original pattern. The other twelve rays will give rise to spurious spots, but (again barring coincidences) these spots will be fainter than the genuine ones. We can therefore expect to be able to pick out the wheat from the chaff with a detector with a threshold slightly less than three units of brightness.

This reasoning applies equally to the "correlograph", with lens, illustrated in Figs. 1 and 2. So, having found how to get rid of the unwanted background in reconstructing A from B and C, we can now envisage the possibility of constructing multiple correlograms, comprising all the spots present in $C_1 = \bar{A}_1 * B_1$ or in $C_2 = \bar{A}_2 * B_2$, and so on. The presentation of B_1 should evoke A_1, presentation of B_2 should evoke A_2, and so on, up to the limit set by the information capacity of the system. But what is this limit ?

To answer this question let us evade the complicated (and basically irrelevant) issues raised by the finite wavelength of light, edge effects and so on, and pose the question in terms of a discrete, and slightly more abstract, model. We suppose A, B and C to be discrete spaces, each containing N points, a_1 to a_N, b_1 to b_N, and c_1 to c_N. The point-pair (a_i, b_j) is mapped on to the point c_k if $i-j=k$ or $k-N$. Conversely, the point-pair (c_k, b_j) is mapped on to a_i if the same condition is met. Imagine now that we have R pairs of patterns which we wish to associate together, each pair consisting of M points selected from A and another M selected from B. The total number of point-pairs determined by all the pairs of patterns will be RN^2, and we may think of this number of "rays" striking C. If they impinge at random, the probability of any point c_k not being struck will be

$$\exp(-RM^2/N) = 1-p, \text{ say}$$

The correlogram for the whole set of R pairs will then consist of the remaining pN points of C.

Now consider the reconstruction process. One of the B-patterns, comprising M of the points b_1 to b_N, is selected, and combined with the correlogram to produce pNM "rays" impinging on A. Each point of the original A-pattern will receive exactly M rays, so that we should set the threshold of our detector at M if we want to pick up all the original points. Now consider any one of the $N-M$ other points in A. It may receive a ray through any one of the M "holes" in B; the probability that it receives a ray through a given hole is just p, for this is the chance that the point on C "behind" the hole belongs to the correlogram. The chance of an unwanted point reaching the threshold is thus p^M, and the probable number of spurious points of brightness M is consequently $(N-M)p^M$. If M is a fairly large number, this will be a sensitive function of p, and for given N and M the critical value of p above which spurious points begin to appear may be found from the relation

$$(N-M)p^M = 1$$

Alternatively, this may be viewed as a relation which sets a lower limit to the value of M for given values of N and p. A slightly safer estimate is given by

$$Np^M = 1, \text{ or } M = -\log N/\log p$$

If M falls below this value, the reconstruction will be marred by spurious points.

Next we enquire about the amount of information stored in the memory when R pairs have been memorized and M satisfies the aforementioned condition for accurate retrieval. We can evoke any one of R A-patterns by presenting the appropriate B-pattern. There are $\binom{N}{M}$ possible A-patterns altogether, so the amount of information needed to store any one of them is $\log \binom{N}{M}$, which is roughly $M \log N$ natural units of information. The total amount of information stored is, therefore, approximately

$$I = RM \log N \text{ natural units}$$

But according to our original calculation of p

$$RM^2 = -N \log (1-p)$$

and if we are working at the limit of accurate retrieval

$$M = -\log N/\log p \simeq \log_2 N \text{ (see below)}$$

It follows immediately that

$$I = N \log p \log (1-p)$$

As one might have anticipated, this expression has its maximum value when p is 0·5—when the correlogram occupies about half of C.

What is remarkable is the size of $I_{\max}$.

$I_{\max} = N(\log 2)^2$ natural units $= N \log 2$ bits. The maximum amount of information that could possibly be stored in C is N bits. So the correlograph, in this discrete realization, stores its information nearly ($\log_e 2 = 69$ per cent) as densely as a random access store with no associative capability.

As described, the discrete correlograph, like the holograph, will "recognize" displaced patterns. If an A-pattern $\{a_i\}$ and a B-pattern $\{b_j\}$ have been associated, then presentation of the displaced B-pattern $\{b_{j+d}\}$ will evoke the displaced A-pattern $\{a_{i+d}\}$.

But the resemblance does not cease there. Just as in holography, the information to be stored is laid down (i) in parallel, (ii) non-locally and (iii) in such a way that it can survive local damage. In parallel, because each mapping $(a_i, b_j) \rightarrow c_k$ can be effected without reference to any other; the same applies to the reconstructive mappings $(c_k, b_j) \rightarrow a_i$. Non-locally, because the presence of a_i in an A-pattern is registered at M separate points on the correlogram, one for each point of the B-pattern. And robustly, because if the system is not stretched to its theoretical limit it can (as we shall show elsewhere) be used for the accurate reconstruction of A-patterns even when some of the correlogram is "ablated" and/or the B-patterns are inaccurately presented. But it can only be made secure against such contingencies by sacrificing storage capacity—as one would expect.

In our discussion of the process of reconstruction we had occasion to note that a point c_k might owe its presence on the correlogram to the joint occurrence of (a_i, b_j); but that if a pattern were presented containing the point b_{j+d}, the "ray" (c_k, b_{j+d}) would light up the point a_{i+d}, which might never have occurred in any A-pattern. It was

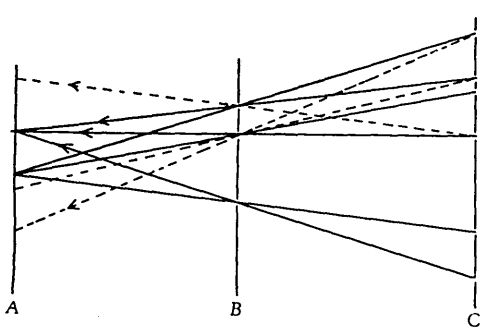

Fig. 3. Showing that original spots are generally brighter.

this feature which underlay the ability of the system to recognize displaced patterns; but the same feature is a slight embarrassment when one comes to consider how a discrete correlograph, with the reconstructive facility, could be realized in neural tissue. We will not dwell on this point, except to acknowledge that it was drawn to our attention by Dr F. H. C. Crick, to whom H. C. L.-H. is indebted for provocative comments. But it led us on to a further refinement of our model, in which a given point c_k is admitted to the correlogram only if the particular pair (a_i,b_j) occurs in one of the pairs of patterns, and not otherwise. On this assumption there might be as many as N^2 separate point-pairs to take into account, and a correspondingly large number of points in the space C.

In this form our associative memory model ceases to be a correlograph, having lost the ability to recognize displaced patterns, but its information capacity is now potentially far greater than before. To show this, we will adopt a rather different type of representation, in which the points of A become N_A parallel lines, and those of B become N_B parallel lines. The points of C are the N_AN_B intersections between the lines a_i and the lines b_j.

In this network model, as before, a particular point of C is included in the active set if the pair of lines (a_i,b_j) which pass through it have been called into play in at least one association of an A-pattern with a B-pattern. Let us suppose that R pairs of patterns have been associated in this way, each pair comprising a selection of M_A lines from A and M_B lines from B. Then the chance that a given point of C has not been activated by the recording is

$$\exp(-RM_AM_B/N_C)=1-p, \text{ say}$$

where we have written N_C for N_AN_B. If B-patterns are being used to recall A-patterns, then there will be a minimum value of M_B such that if the threshold on the A-lines is set at M_B (so as to detect all the genuine lines) spurious lines will begin to be detected as well. (The argument is just the same as that applied to the correlograph earlier on.) This minimum value of M_B is given by

$$N_Ap^{M_B}=1$$

or
$$M_B=-\log N_A/\log p\simeq\log_2 N_A$$

Now the amount of information stored in the memory when R pairs of A-patterns have been memorized is roughly

$$I_A=RM_A \log N_A$$

But from our equation for $1-p$

$$RM_AM_B=-N_C \log (1-p)$$

therefore

$$I_A=N_C \log p \log (1-p)$$

showing that, as in the correlograph, the density with which the associative net stores information is 69 per cent of the theoretical maximum value. We may note, in passing, that I_B, defined as $RM_B \log N_B$, is also equal to $N_C \log p \log (1-p)$.

An associative network of this kind also operates (i) in parallel (ii) non-locally and (iii) in such a way that local damage or inaccuracy is not necessarily disastrous. We intend to go into the details of (iii) elsewhere. We now succumb to the temptation of indicating how such an associative memory might be realized in neural tissue though, as Brindley has pointed out[6], function need not determine structure uniquely.

The system we have in mind is represented diagrammatically in Fig. 4. The horizontal lines are axons of the N_B input neurones $b_1,b_2, \ldots$, while the vertical lines are dendrites of the N_A output neurones $a_1, a_2, \ldots, \ldots$. At the intersection of b_j with a_i is a modifiable synapse c_{ij}. This synapse is initially inactive, but becomes active after a coincidence in which a_i and b_j are made to fire at the same time by some external stimulus. Such a coincidence is

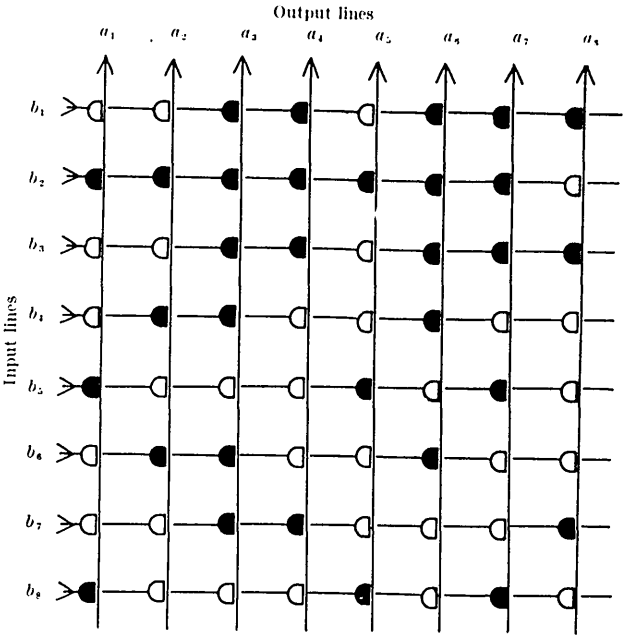

Fig. 4. An associative net.

supposed to occur if an A-pattern containing a_i is presented in association with a B-pattern containing b_j. After the activation of c_{ij} (which we regard as a permanent effect) the firing of b_j will locally depolarize the membrane of a_i. The output neurone a_i is then supposed to fire if M_B or more input cells depolarize it simultaneously.

In Fig. 4 we indicate what the state of the network would be after it had learned to associate the following pairs of patterns:

B-pattern	A-pattern
1,2,3	4,6,7
2,5,8	1,5,7
2,4,6	2,3,6
1,3,7	3,4,8

The synapses indicated by solid semicircles would be active, those indicated by open semicircles being still inactive. In this particular example, N_A and N_B are both 8, and M_A ($\simeq\log_2 N_B$) and M_B ($\simeq\log_2 N_A$) are both 3. R, the number of pairs of patterns associated, has been chosen so as to make p, the proportion of synapses active, close to 0·5; in fact p equals 0·5 exactly. These various numbers illustrate the system working near its maximum capacity. The reader may verify that every B-pattern except the first evokes the correct A-pattern at a threshold of 3; the only mistake the system makes is that when supplied with the B-pattern 1,2,3 it responds with an A-pattern 3,4,6,7 containing four elements.

To summarize, we have attempted to distil from holography the features which commend it as a model of associative memory, and have found that the performance of a holograph can be mimicked and actually improved on by discrete non-linear models, namely the correlograph and the associative net just described. Quite possibly there is no system in the brain which corresponds exactly to our hypothetical neural network; but we do attach importance to the principle on which it works and the quantitative relations which we have shown must hold if such a system is to perform, as it can, with high efficiency.

Received March 17, 1969.

[1] Van Heeden, P. J., *App. Optics*, **2**, 393 (1963).
[2] Pribram, K. H., *Sci. Amer.*, **220**, 73 (1969).
[3] Willshaw, D., and Longuet-Higgins, H. C., *Machine Intelligence* 4 (edit. by Michie, D.) (Edinburgh University Press, 1969).
[4] Gabor, D., *Nature*, **217**, 1288 (1968).
[5] Gabor, D., *Nature*, **217**, 584 (1968).
[6] Brindley, G. S., *Proc. Roy. Soc.*, B, **168**, 361 (1967).

From Machine Intelligence 5

Edinburgh University Press

1970

19

Associative Memory Models

D. J. Willshaw and
H. C. Longuet-Higgins
Department of Machine Intelligence and Perception
University of Edinburgh

1. INTRODUCTION

In a recent article in *Nature* written in collaboration with O. P. Buneman (Willshaw, Buneman and Longuet-Higgins, 1969) we described two quasi-holographic devices for the associative storage and retrieval of complex patterns. These devices, the correlograph and the associative net, serve many of the same purposes as the holograph, but are simpler in conception and lend themselves more easily to computer simulation. Our interest in them is twofold: first, it is quite possible that the principles on which they work are employed in the central nervous system; and secondly, they may well find application in computing technology, especially when parallel computation techniques become generally available.

2. THE OPTICAL CORRELOGRAPH

The correlograph was originally envisaged as an optical analogue device, illustrated in figures 1 and 2. It is designed for storing the correlation C between a pair of patterns A and B in such a way that A can be retrieved from B and C, or B retrieved from A and C. A and B are patterns of pinholes through black cards. When A is illuminated from the left, rays pass through the holes in A and B and are guided by a lens L on to a screen at C. (The distances AB and LC are both set equal to the focal length of L.) A 'correlogram' is then made by taking a black card and making a pinhole through it at every point at which a ray strikes the viewing screen. This correlogram is mounted at C and illuminated from the right, so that rays now pass through the holes in C and the holes in B. A pattern of bright spots now appears at A, and the brightest of these spots coincide with the pinholes of the original pattern A. A can therefore be reconstructed by sifting out the brightest spots with a threshold detector. (Also, though this need not concern us, B can be reconstructed by turning C upside down, placing A where B was before, illuminating from the

right and mounting the threshold detector where A was before. The reader may care to satisfy himself of the truth of this assertion.)

A further use of the correlograph is for detecting the presence of a particular configuration of pinholes in the pattern B. In this application the pinhole

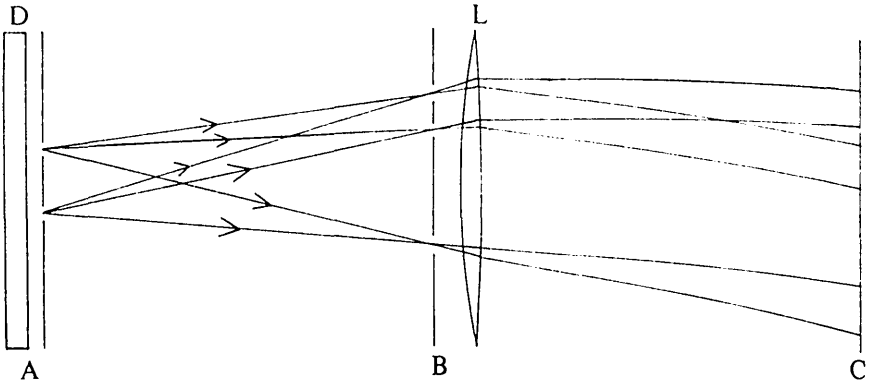

Figure 1. Constructing a correlogram. D is a diffuse light source, L a lens and C the plane of the correlogram of A with B

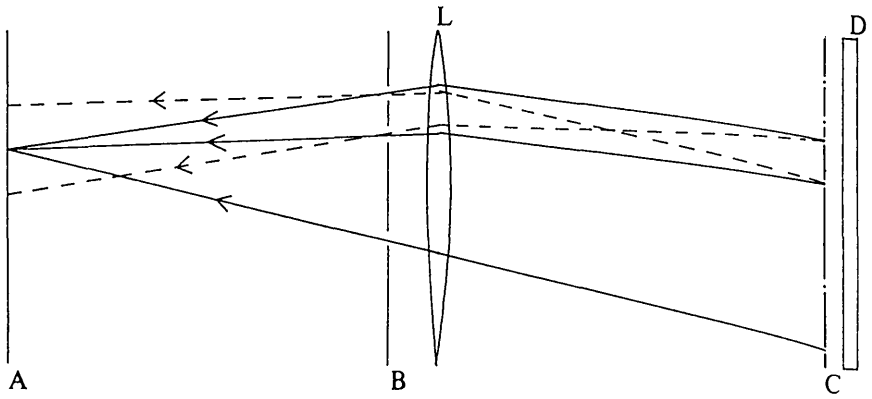

Figure 2. Reconstructing a pattern. Full lines are paths traversed in Figure 1. Broken lines are paths *not* traversed in Figure 1

pattern on A is a copy of the configuration to be looked for, and an exceptionally bright spot will appear on C at every point which lies opposite to an occurrence of A in B. Furthermore – and this is something that cannot be done holographically – if a scaled-down version of A appears somewhere in B, this fact can be detected by moving the viewing screen towards the lens L, when an exceptionally bright spot will appear on the screen at an appropriate distance from L. The correlograph can therefore be used as a pattern-recognition device possessing both displacement invariance and size invariance,

But the main interest of the device is its ability to store simultaneously the correlations between several different pairs of patterns, in such a way that either member of any pair can be used for reconstructing the other. Let C_1 be the correlogram between A_1 and B_1, let C_2 be the correlogram between A_2 and B_2, and so on. Then (up to a certain limit of saturation, which we shall discuss below) one can achieve this multiple performance by constructing a joint correlogram C which is the union of the various pinhole sets C_1, C_2, etc. For the optical correlograph the exact theory of this application is difficult to formulate precisely, because to obtain a high information storage density the pinholes must be small; but if they are too small diffraction effects seriously complicate the situation, which can no longer be described by geometrical optics. A further complication is presented by edge effects in an optical realization of the correlograph. We shall therefore not discuss this realization further, but proceed at once to a consideration of the digital correlograph, in which both these complications are avoided.

3. THE DIGITAL CORRELOGRAPH

The digital correlograph is an abstract system in which the cards A and B of the optical correlograph are regarded as discrete spaces each comprising N points. An A-pattern or a B-pattern is a choice of M points from one of these spaces. The product space of A and B is mapped on to a third space C, which also comprises N points, in the following systematic manner: the point pair (a_i, b_j) is mapped on to the point c_k if and only if $j-i=k$ or $k-N$. (Each of the three subscripts runs from 1 to N.) There is a converse mapping of point-pairs from C and B on to points of A: (c_k, b_j) is mapped on to a_i if and only if $j-k=i$ or $i-N$. This converse mapping is employed in the reconstruction of an A-pattern from the correlogram and the paired B-pattern. Another converse mapping allows the retrieval of a B-pattern from its A-pattern: the point pair (c_k, a_i) is mapped on to b_j if and only if $i+k=j$ or $j-N$. (The + sign in the last equation may serve as a hint to the reader who has not yet solved the exercise offered in the preceding section.)

Now let us consider briefly the storage of several pairs of patterns. Each pair maps onto M^2 of the N points of C, but these M^2 points may not all be distinct. After the association of R pairs of patterns, RM^2 (not necessarily distinct) points on C will have been added to the correlogram. Let pN be the number of distinct points of C which belong to the multiple correlogram. Then if the recorded patterns are random and uncorrelated we may safely assume that

$$1-p=\exp(-RM^2/N). \tag{1}$$

In the reconstruction of an A-pattern from the correlogram and the associated B-pattern we employ the converse mapping $(c_k, b_j)\rightarrow a_i$, for all points c_k belonging to the correlogram and all points b_j of the B-pattern. A particular point belonging to the original A-pattern will be activated M times; but the

chance that a point not belonging to the A-pattern will be activated M times is only p^M. There are N points altogether in the space A; so we can be fairly sure of not obtaining any spurious points at a threshold of M if

$$Np^M < 1. \tag{2}$$

Regarding the latter equation as an equality defining the point of saturation, we may rewrite (1) and (2) as

$$RM = -(N/M)\log_e(1-p) \tag{3}$$

and $$\log_e N = -M \log_e p \tag{4}$$

respectively. From these two equations it is an easy matter to determine the density at which information is stored in C when the digital correlograph is working near saturation. The information content of a single A-pattern is, in natural units rather than bits,

$$\log_e \binom{N}{M};$$

so when R such patterns can be retrieved with accuracy the stored information amounts to

$$I_e = R \log_e \binom{N}{M} = RM \log_e N \text{ approximately}$$

(provided M/N is small, as we shall soon verify). So by (3) and (4)

$$I_e = N \log_e p \log_e (1-p). \tag{5}$$

This expression is obviously maximal when $p = \frac{1}{2}$; so the maximum number of natural units which can be reliably retrieved is $N(\log_e 2)^2$ natural units. Noting that 1 bit $= \log_e 2$ natural units, and that $\log_e 2 = 0{\cdot}69$, we conclude that we can store information in the space C, regarded as a set of binary registers, to a density of 0·69 bits per register — nearly 70 per cent as densely as information is stored in a random access store!

To complete this discussion we note that when the system is being stretched to its limit, so that $p = \frac{1}{2}$, equation (4) implies that $M = \log_2 N$, so that M/N is rather small, as assumed in approximating I; and that the number of pairs of patterns which have been associated is given by $R = N \log_e 2/(\log_2 N)^2$.

4. A COMPUTER SIMULATION

We have tested our theoretical results by computation, taking $N = 256$ and $M = \log_2 N = 8$. The theoretical value for R at saturation is $4 \times 0{\cdot}693 = 2{\cdot}77$, so we should be able to store two pairs of patterns with accurate retrieval, and three pairs with only slightly inaccurate retrieval. Such is indeed the case. With one pair of random 8-point patterns the correlogram is found to comprise 59 points, and retrieval is perfect. When a second pair is loaded the number of points of C involved rises to 101, and either member of each pair can still be retrieved without error from the other. When a third pair is loaded, the correlogram comprises 136 of the 256 points of C. Patterns A_1

and B_1 still retrieve one another without error, but inaccuracies arise in the mutual retrieval of A_2 and B_2, and of A_3 and B_3. In the reconstructions of A_2, B_2, A_3 and B_3 the numbers of spurious points appearing are 1, 1, 2 and 0 respectively; so we are beginning to witness breakdown, and this becomes catastrophic if any more pairs are loaded into the system.

With random 5-point patterns the system works less efficiently; it can only be loaded to a p value of about 0·3 without the appearance of many spurious points in the retrieved pattern:

Number of stored pairs	Number of points in correlogram	Divided by 256	Mean number of spurious points
1	25	0·1	0
2	49	0·2	0
3	69		0
4	85	0·3	0·25
5	104	0·4	1·4
6	121		3·3
7	131	0·5	7

5. THE ASSOCIATIVE NET

The associative net is logically similar to the digital correlograph, and performs much the same function, but it sacrifices the displacement invariance of the correlograph in return for a much greater absolute storage capacity. As shown in figure 3 the information is stored in the associative net in a set of on/off switches which are situated at the $N_A N_B$ intersections between a set of N_A A-lines and a set of N_B B-lines. A pair of patterns is stored associatively by sending pulses down M_A chosen A-lines and at the same time sending pulses down M_B chosen B-lines, and turning on switch c_{ij} if the lines a_i and b_j both carry pulses. A second pair of patterns is stored by repeating this process and turning on further switches by the same rule; if such a switch is already on, it is simply left on (but see later). Retrieval of an A-pattern from a B-pattern is effected by sending M_B pulses down the appropriate B-lines, and these are transmitted to certain A-lines through the switches which happen to be on. Each A-line is fitted with a threshold detector, and if the line receives as many as M_B pulses through its switches it discharges a pulse. The A-lines which discharge pulses will include those of the required A-pattern; one will try to arrange that no other lines also discharge.

In the absence of damage, or inaccuracy in the cue pattern, the theory goes very much as for the digital correlograph. The corresponding relationships are:

$$RM_A = -(N_A N_B / M_B) \log_e (1-p),$$

$$\log_e N_A = -M_B \log_e p,$$

and
$$I = R \log_e \binom{N_A}{M_A} = RM_A \log_e N_A = N_A N_B \log_e p \,.\, \log_e (1-p).$$

Again we find that the maximum information density is about 0·69 bits per switch for a large net, and we conclude that in order to attain this maximum density we must have $M_B = \log_2 N_A$.

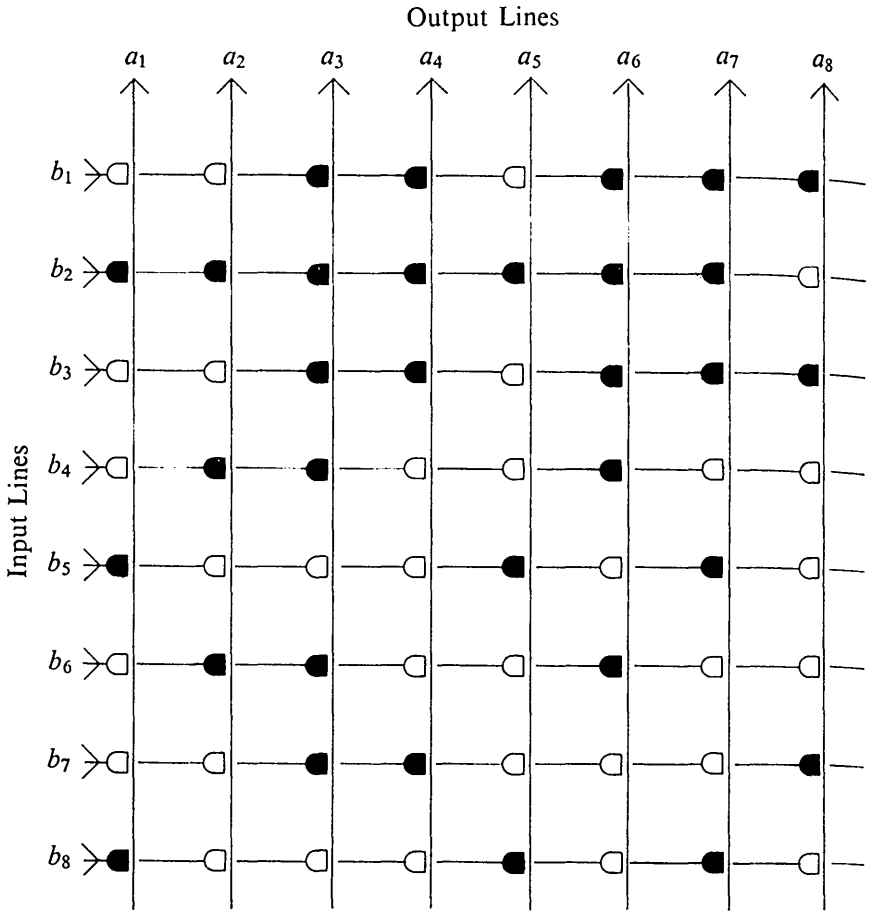

Figure 3. Associative net

The critical value of R, the number of stored patterns, is then

$$R = (N_A N_B / M_A M_B) \log_e 2.$$

If the system is being used both for retrieving As from Bs and vice versa, optimal efficiency requires also that $M_A = \log_2 N_B$ and R takes the alternative form $R = N_A N_B \log_e 2 / (\log_2 N_A \,.\, \log_2 N_B)$.

For example, with $N_A = 10^6 = N_B$ (figures which might not be unreasonable for a piece of cerebral cortex) we should be able to associate approximately $1{\cdot}5 \times 10^9$ pairs of 20-impulse patterns before saturating the net.

6. THE EFFECTS OF DAMAGE AND INACCURACY

Since the correlograph and the associative net store information in a distributed manner, one might expect them to be able to perform well in the face of damage or inaccurately presented cues. We will therefore imagine that initially the associative net was loaded so that a fraction p of the switches were turned on, but that some disaster then occurred which caused a fraction $1-q$ of these switches to be turned off, at random, so that only a fraction pq remains on.

Consider now what happens when a stored B-pattern is fed into the B-lines. An A-line which ought to put out a pulse will only receive qM_B pulses on the average, rather than M_B, so if this line is to fire the threshold must be lowered to tM_B, say, where t is sufficiently smaller than q. But t must not be too small, or else the A-lines which ought to be silent will emit pulses; these A-lines will receive pqM_B pulses on the average, so t must be safely larger than pq. Straightforward statistical analysis shows, in fact, that the evoked A-pattern will be accurate only if the following two inequalities hold:

$$(1/M_B)\log_e M_A < t\log_e(t/q) + (1-t)\log_e((1-t)/(1-q)) \quad (6)$$

$$(1/M_B)\log_e N_A < t\log_e(t/pq) + (1-t)\log_e((1-t)/(1-pq)) \quad (7)$$

The former inequality must be satisfied if all the A-lines which ought to fire are to do so; the latter if all the others are to remain silent.

We have made some calculations on the performance of the damaged associative net for which $N_A = N_B = N = 10^6$, and $M_A = M_B = M$. We began by choosing a value for q in the range 0·4 to 1·0, and used equation (6), regarded as an equality, to find a functional relation between M and t. Next we used (7), also regarded as an equality, to find a functional relation between p and M, which is equivalent to a relation between p and t. For given t, p and q the fractional capacity of the net, in bits per switch, can be calculated as follows:

$$I_2 = R\log_2\binom{N}{M} = RM\log_2(N/M) \quad (8)$$

gives the amount of stored information, in bits, to a rather better approximation than that used in sections 3 and 5. But

$$R = -(N^2/M^2)\log_e(1-p). \quad (9)$$

Therefore the mean number of bits per switch is, by (6), (7), (8) and (9),

$$I_2/N^2 = \log_2(1-p)(t\log_e p + (1-t)\log_e(1-pq)/(1-q)). \quad (10)$$

For given q and t this expression has a maximum when p is between 0 and 0·5, and this maximum is easily calculated. So there are definite values of

t, M and p which optimize the performance of the system for given q, and these are given in the following table:

$q=$	0·4	0·5	0·75	0·9	0·95	1·00
$t=$	0·261	0·340	0·557	0·731	0·811	1
$p=$	0·224	0·24	0·257	0·30	0·337	0·5
$M=$	111	84	43	31	27	20
$R=$	$2{\cdot}1\times10^{7}$	$3{\cdot}9\times10^{7}$	$1{\cdot}6\times10^{8}$	$3{\cdot}7\times10^{8}$	$5{\cdot}6\times10^{8}$	$1{\cdot}7\times10^{9}$
$I/N^2=$	0·03	0·044	0·10	0·18	0·23	0·54

The fact that the last figure, 0·54, is rather less than $\log_e 2$, is due to the finiteness of the net. The corresponding figures for the damaged correlograph would be exactly the same except for the R values, which would all be smaller by a factor of 10^6.

The main conclusion to be drawn from these figures is that although damage reduces seriously the stored information density, the associative net and the correlograph can still be made to work reliably under such adverse conditions by increasing the number of pulses in the input patterns and being careful not to load too many pairs of patterns into the system. A further point which should be noted is that precisely the same calculations apply to a situation in which there is no damage, but the input cues are incomplete. The absence of a pulse in an input line which ought to carry one is logically equivalent to a malfunctioning of a switch which ought to be on but is in fact off. So the above table applies to the case in which q represents not damage to the switches but the degree of completeness of the input pattern – the number of input pulses being not M but only qM. So again, if the value of M is made large enough, and not too many pairs are loaded into the system, we can obtain accurate retrieval of A-patterns from incomplete B-patterns.

7. LEARNING WITH FORGETTING

So far, in considering the digital correlograph and the associative net, we have assumed that every switch which is turned on remains on unless it is turned off by accidental damage. We have found that in the absence of damage there is a fairly sharp limit to the number of associations that can be stored; if this number is exceeded the performance of the system degenerates rapidly. But is it not possible to store an indefinite number of associations, if one is prepared to forget old ones gradually as new ones are recorded? We now consider this problem.

Let us assume, then, that an associative net has been loaded with a number of pairs of patterns, and that pN^2 of its N^2 switches are now on. A new pair of patterns is to be stored, but the value of p is not to increase. The new pair of patterns will call for the turning on of M^2 switches, but pM^2 of these will already be on. If each of the others is turned on with probability s, then the

total number of switches turned on in recording the new pair will be $s(1-p)M^2$. The total number of switches which are already on is pN^2, so to maintain p at its previous value we must turn off each of these with probability r given by

$$rpN^2=s(1-p)M^2.$$

After a new association has been recorded, then, each of the relevant switches will be on with a probability $p+s(1-p)$, but this probability will fall steadily towards p as other associations are recorded. Suppose that at any stage it has a probability $p+z$ of still being on. Then after one more recording this chance becomes $p+z'$, where

$$p+z'=(1-r)(p+z)+(sM^2/N^2)(1-p-z).$$

Using the fact that $r=s(1-p)M^2/pN^2$ we deduce straightforwardly that

$$z'=z(1-sM^2/pN^2).$$

Remembering that M^2/N^2 is small in general, we rewrite this as

$$z'=z\exp(-sM^2/pN^2),$$

and deduce that after the recording of n other associations the value of z will have dropped by a factor $\exp(-nsM^2/pN^2)$. An alternative form for this factor may be derived in terms of R, the effective number of stored patterns, namely

$$R=-(N^2/M^2)\log_e(1-p)=pN^2/M^2 \text{ approximately.}$$

So if the original association is overlaid with n other associations, all of the same strength s, the final value of z will be given by

$$z^{(n)}=z^{(o)}\exp(-ns/R).$$

By an obvious generalization, if the subsequent associations have strengths $s_1, s_2, \ldots, s_n$, then the exponential factor on the left becomes

$$\exp-(s_1+s_2+\ldots+s_n)/R.$$

In particular, if each recording is of full strength, the memory length is just R, which is the effective number of associations in store.

In order to make effective use of an associative net which forgets earlier messages slowly as it learns new ones, it will of course be essential for the threshold on each output line to be less than the M value of any input signal which is to survive being overlaid by a substantial number of other signals. In fact, the effect on a recorded signal of overlaying it with others is much the same as the effect of damage, discussed in section 6. There are two ways of increasing the survival time of a particular association: either recording it several times in succession – 'overlearning' it – which has the effect of raising its effective strength to near unity; or increasing the number M of pulses in the input pattern, so that it takes a longer time for the number of effective pulses to fall below the preset threshold. The mathematical analysis of this situation would, however, take us beyond the scope of this paper.

REFERENCE

Willshaw, D.J., Buneman, O.P. and Longuet-Higgins, H.C. (1969) *Nature*, **222**, 960.

(*Reprinted from Nature, Vol. 225, No. 5228, pp. 177–178, January 10, 1970*)

Models for the Brain

Willshaw, Buneman and Longuet-Higgins have proposed a nonholographic associative memory model for the brain[1]. They also criticize the proposal made by myself[2] and by Pribram[3,4] that the brain would be organized on the holographic principle. They say: "How could the brain Fourier-analyse the incoming signals with sufficient accuracy, and how could it improve on the rather feeble signal to noise ratio of the reconstructed signals ?".

In an earlier paper[5], in which the potential of the hologram for retrieving information was first pointed out, I calculated the signal to noise ratio. As an example I showed that, theoretically, in a hologram of a library of 300 books in coded form of 200 pages each, one single line could instantaneously be recognized and located. The hologram contains half the information held in the ideal matched filter. One can show, for example, that two holograms, in the two arms of a Michelson interferometer, perform the same function as one matched filter. Further, the hologram of a symmetric image, which has half the information of a general image, is identical with the matched filter[5]. It therefore seems correct to say that the signal to noise ratio of the hologram is 50 per cent of ideal. No method of information retrieval can have a signal to noise ratio better than the hologram by more than a factor of two.

In a book on the subject[6] I discussed further how the brain could work physically very well as a three-dimensional hologram. If we have a three-dimensional network of neurones, in which each neurone is connected to a few adjacent ones, and if a neurone in a certain layer, in receiving a signal, will send this on to a few neurones in the next layer, then signals will propagate in this network as a wave propagates in an elastic medium. If, moreover, the ability of the neurones to propagate received signals can be permanently enhanced by frequent use, then the network must act as a three-dimensional hologram, with a storage capacity of the order of the number of neurones present in the network.

For recognizing, we need a two-dimensional hologram for fast search, combined with a three-dimensional hologram which has a large capacity for storing information that is readily accessible[6]. This is still not sufficient, however, to explain the wonderful human capacity for recognizing. We can recognize a person, even one we have not met for a long time, at any distance and from many

different angles. A fixed hologram memory would not be able to perform this operation. The flexibility needed can be provided by optical means: for example, a zoom lens can carry out a search to match the size of the image received to the image stored. It seems not too far-fetched to imagine that a neurone network has this flexibility. It could be realized by extended variable fields, analogous to those used in electron optics, to produce different gradients in the speed of propagation of the network by electrical or chemical means. This could effect a change in focal distance, or a rotation of the image, or small distortions, to achieve a clear, sharp recognition signal in the image plane.

Although the hologram principle is natural for a neurone network, it does not exclude the possibility that another model such as the correlogram of Willshaw, Buneman and Longuet-Higgins is actually realized in the brain. One has first, however, to show that such a model is reasonable. Their model, in the optical form they propose, seems to have a low storage capacity because of the diffraction of any kind of wave field (this is not irrelevant!). In the network model they propose, on the other hand, they do obtain the same storage capacity as the holographic model, but it seems to lack the flexibility for recognizing images which are displaced, of different size, or slightly distorted. One more aspect to be considered is the fact that three-dimensional holograms are capable of storing time dependent signals[2]. The recognition of speech, and our ability to speak or run or drive a car, is one more aspect of information processing in the brain which must be explained by any model.

P. J. VAN HEERDEN

Polaroid Research Laboratories,
Cambridge, Massachusetts 02139.

Received October 20, 1969.

[1] Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C., *Nature*, 222, 960 (1969).
[2] van Heerden, P. J., *Applied Optics*, 2, 393 (1963).
[3] Pribram, K. H., in *Macromolecules and Behavior* (edit. by Gaito, J.) (Academic Press, New York, 1966).
[4] Pribram, K. H., *Sci. Amer.*, 220, 73 (1969).
[5] van Heerden, P. J., *Applied Optics*, 2, 387 (1963).
[6] van Heerden, P. J., *The Foundation of Empirical Knowledge, with a Theory of Artificial Intelligence* (Wistik, Wassenaar, Netherlands, 1968).

VAN HEERDEN has discussed some of the differences between his holographic model of memory[1] and a pair of non-holographic models that we put forward last year[2]. We alluded to the poor signal to noise ratio of the holograph when it is used to reconstruct a stored pattern from a fragment of that pattern; van Heerden's comment refers

to the signal to noise ratio with which one can locate a given fragment in a large text, and this is quite a different matter. Van Heerden himself showed that the signal to noise ratio in the reconstruction of random patterns was equal to the size of the fragment divided by the size of the whole pattern, and the same applies to the reconstruction of a temporal signal from a short cue[3].

Our first model, the correlograph, was designed to re-create accurately one binary pattern from another, having recorded a cross-correlation between the two. To do this with accuracy it was necessary to put a threshold on the output of the device, so that the relatively weak unwanted signals would not appear in the reconstruction. Regarding the output pattern or patterns (several input–output pairs can be stored simultaneously) as constituting the stored information, we found that the information storage density could be as high as 69 per cent of the theoretical maximum without loss of accuracy in recall.

These remarks apply equally to our other model, the associative net. Although more information can be stored in the associative net, we lose the ability to produce a displaced output from a correspondingly displaced input. In this and in some other respects the net behaves like a discrete version of van Heerden's three-dimensional hologram[4], but, again, threshold elements are used to clean up the output patterns, a result which cannot be achieved in a purely linear device.

In its optical form the correlograph is, indeed, severely limited by diffraction and cannot be taken literally as a physical model for memory; nor did we intend that it should. But we felt that its logic, which is easy to appreciate, might possibly be realized in the nervous system. For instance, an associative net might be made to function as a correlograph by "tying together" certain of its switches. But there was no evidence that any such tying takes place and we therefore put forward the associative net as the more likely model. It could be simply realized, as we pointed out, by a system of neurones with thresholds and modifiable synapses; both these properties are known to occur peripherally in the nervous system[5,6] and probably occur centrally as well[7].

Although there is no conclusive neurophysiological evidence to support our theory against van Heerden's, the ability of parts of the nervous system to propagate waves according to Huygens's principle would be difficult to reconcile with the observed non-linearity of some neural responses, and the existence of a stable periodic source of excitation has yet to be demonstrated. We also feel that in any model of the brain it is of advantage to be able to modify synapses as well as nerve cells. The ratio of synapses to nerve cells in the cerebral cortex seems to be of the order

10^4–10^5 so that the information that could be stored synaptically would be correspondingly higher[8]. As to the remarkable flexibility of the human perceptual apparatus, we feel that neither his model nor ours can be held to account for this in their present forms.

D. J. WILLSHAW
H. C. LONGUET-HIGGINS
O. P. BUNEMAN

Department of Machine Intelligence and Perception,
University of Edinburgh.

Received November 10, 1969.

[1] van Heerden, P. J., *Applied Optics*, 2, 387 (1963).
[2] Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C., *Nature*, 222, 960 (1969).
[3] Willshaw, D. J., and Longuet-Higgins, H. C., in *Machine Intelligence*, 4 (edit. by Michie, D.) (Edinburgh Univ. Press, 1969).
[4] van Heerden, P. J., *Applied Optics*, 2, 393 (1963).
[5] Eccles, J. C., *The Physiology of Nerve Cells* (Johns Hopkins Press, Baltimore, 1957).
[6] Eccles, J. C., *The Physiology of Synapses* (Springer, Berlin, 1964).
[7] Burns, B. D., Bliss, T. V. P., and Uttley, A. M., *J. Physiol.*, 195, 339 (1968).
[8] Cragg, B. G., *J. Anat.*, 101, 639 (1967).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.

Quarterly Reviews of Biophysics, 3, 2 (1970), pp. 223–244
Printed in Great Britain

Theories of associative recall

H. C. LONGUET-HIGGINS, D. J. WILLSHAW AND O. P. BUNEMAN
Department of Machine Intelligence and Perception, University of Edinburgh, 2 Buccleuch Place, Edinburgh

1. INTRODUCTION

The problem of how the brain stores and retrieves information is ultimately an experimental one, and its solution will doubtless call for the combined resources of psychology, physiology and molecular biology. But it is also a problem of great theoretical sophistication; and one of the major tasks confronting the brain scientist is the construction of theoretical models which are worthy of, and open to, experimental test. In this review we shall be concerned with the latter aspect of the problem of memory, which has attracted quite a lot of attention in the last few years. It is early yet to judge the relative merits of the various models in any detail; but as we shall see, most of those which have been developed

beyond their initial hypotheses have a certain family resemblance, and it seems as if we may now be in possession of the basic ideas which will be needed for the understanding of one of the central problems of memory, namely the mechanism of associative recall.

The most convenient definition of associative recall, for our purposes, is in terms of stimuli and responses. Consider a physical system occupying a certain region of space—a 'black box', in fact. The box has an input channel and an output channel, each capable of transmitting very complex signals. The input signals are of two kinds, which we may designate as conditioned (CS) and unconditioned (UCS) stimuli respectively. Before the system has memorized anything, any UCS will evoke a certain response which will emerge along the output channel. During the learning phase the system is subjected to a succession of UCSs, each of which is accompanied by a certain CS, simultaneously or nearly so. After learning is completed the input of a CS will evoke the same response as the UCS which accompanied it during learning; the system will respond to the conditioned stimuli in the absence of the unconditioned stimuli.

A very simple system of just this kind was postulated by Hebb (1949) who advanced a famous hypothesis about the modification of synaptic connections between neurons. The axons of two neurons A and B are connected to the body or the dendrites of a third cell C. The synapse BC is unmodifiable, and an impulse from B will invariably excite C. The synapse AC is initially ineffective; but if C is fired by B at the moment when an impulse arrives from A, then the synapse AC is facilitated, so that thereafter an impulse from A may suffice to excite C without the assistance of an impulse from B. The cell C thus learns to respond to the CS from A in the same manner as it originally responded only to the UCS from B.

There are certain difficulties, however, in generalizing this simple system into an acceptable theory of associative recall. First, an associative memory must be able to associate signals which are highly unexpected, in the sense of having a low prior probability, and it would be extravagant to reserve a separate cell for every conceivable input signal. And even if this were not so, the memory would be unable to establish arbitrary associations unless there existed at least one cell C connected from (i.e. postsynaptically connected with) every possible pair of cells A and B. These remarks underline the desirability of constructing theoretical models in which, after the establishment of many arbitrary associations, the relevant information will have been

stored sufficiently densely not to waste too many modifiable elements. Another generally desirable feature is that the modifications which result from a particular association should be fairly widely distributed over the system, so that the recall of a particular response is not too sensitive to local damage, or to inaccuracy in the CS which should evoke it. And finally, though this is still a largely unsolved problem, the system must be able to associate together signals which are temporally as well as spatially complex, and this calls for components with characteristic frequencies or time constants.

In the following paragraphs we shall not attempt a complete survey of all the published literature on learning in the nervous system, but will concentrate on the rather few papers which have dealt specifically with the neurological problem just posed. The ideas and models which we shall discuss are scattered fairly widely in the literature, in such diverse fields as neuroanatomy, the technology of pattern recognition, and applied optics. We will begin with the optical and quasi-optical models.

2. Optical models

The forerunner of most of the current optical models was a paper by Beurle (1956), though Beurle's ideas were couched not in optical but in neurological terms. Like Cragg & Temperley (1954) and later Griffith (1963, 1965), Beurle explored the hypothesis that the seat of the memory is macroscopically homogenous, such structure as it possesses being adequately describable in terms of local neural connections. This assumption implies that its behaviour can be adequately specified by a set of differential equations, expressing the manner in which it transmits waves of excitation and the manner in which the local parameters of the equations are altered after the passage of such waves. The mathematical details of Beurle's paper need not detain us, as the underlying assumptions are difficult to reconcile with the known structural complexity of the cortex; but one particularly attractive idea emerged from Beurle's analysis, namely that two different waves spreading across the cortex might together generate an interference pattern from which either wave alone could subsequently regenerate the other. Broadly enough interpreted this suggestion could hardly be doubted, and it was explicitly referred to by van Heerden (1963*a*, *b*) in a pair of papers which first took seriously the analogy between associative memory and the optical technique of holography.

Holography was invented by Gabor in 1948 (see also Gabor 1949, 1951) for recording photographically an interference pattern between two light waves in such a way that either light wave, when falling on the interference pattern, regenerates the other. For this purpose spatially coherent light is required, and it was not until the advent of the laser, and the technical developments made by Leith & Upatnieks (1962), Stroke (1966) and others that holography became a practically useful technique for the storage and retrieval of information. The best known holographic experiment is that in which the beam from a laser is split into two by a half-silvered mirror. One part, the reference beam B, is shone directly on to a photographic plate; the other is used to illuminate an object in such a way that the light wave A scattered by the object also falls on the plate. The plate records an interference pattern, and is developed and printed as a positive transparency. When this transparency—the 'hologram'—is now illuminated by the reference beam B, the light wave A is regenerated, and an observer looking through the hologram sees a clear image of the original object, as through a window. Remarkably enough, this image is three-dimensional, and can be seen through any part of the hologram, indicating that every point on the object affects every point on the plate, and conversely that every point on the plate records something about the object as a whole. In both these senses the information storage is non-local, or distributed in the sense of the previous section, and it is probably this fact which has most strongly commended holography to some neurologists as a memory model (Westlake, 1967; Pribram, 1966, 1969).

But there is another feature of holographic recording which is possibly of greater theoretical interest, and it is rather less well known. This is the possibility of storing many different associations on the same hologram. To do this one replaces the direct reference beam of the previous experiment by the light scattered by a second object, so that the plate records the interference pattern between two scattered waves (Stroke, 1966). Either scattered wave will then regenerate the other, though with some loss of definition. But before developing the plate one can expose it to many different pairs of scattered waves, A_1 and B_1, A_2 and B_2, and so on. Then, provided that each pair of scattered waves is effectively random, one can recover the image of any recorded object by illuminating the hologram with the laser light scattered by its partner (which must be placed accurately in its original position). The more pairs of objects that are associated in this way, the worse the definition in the recon-

structed image; but this fact merely exhibits the finite information storage capacity of any physical system—a limitation from which the brain can hardly be exempt.

The theory of the two-dimensional hologram was worked out in detail by van Heerden (1963*a*). He calculated the signal-to-nose ratio of the reconstructed image, and showed how the phenomenon of ghost images could be used for recognizing and locating a small fragment in a larger pattern, using an apparatus of the type described in the next paragraph. In a later paper (van Heerden, 1963*b*) he showed theoretically that optical information could also be stored in light-sensitive three-dimensional systems such as discoloured crystals. The information is stored, as in the two-dimensional hologram, as a series of interference patterns between pairs of plane parallel waves. Many different pictures could be stored in the same crystal, if each was illuminated by a different plane wave, and the number of bits of retrievable information was comparable to the number of colour centres in the crystal, so that the system could be regarded as rather efficient. Relating these results to the suggestions of Beurle (1956), van Heerden stressed the need for exact phase relations between the waves to be maintained over large distances. He postulated a calibrating system of pulses to compensate for any variation in the speed of propagation, and suggested that the memory may comprise two subsystems, one for search and recognition and the other for actual storage. But though his model is mathematically attractive, its physical assumptions are rather speculative in relation to the brain, so we will pass on to consider the underlying theory, which does seem to have more relevance to our neurological problem.

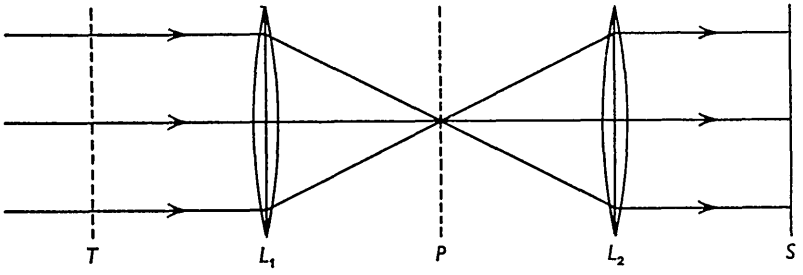

Fig. 1. A Fourier holograph. T is a transparency, illuminated from the left by a collimated laser beam; L_1 and L_2 are convex lenses of focal length d, separated by a distance $2d$; P is a plate holder for the hologram and S is a viewing screen.

Perhaps the easiest holographic experiment to analyse theoretically is Fourier holography, illustrated in Fig. 1.

T is a transparency, illuminated from the left by a collimated laser beam, P is a plate holder and S a ground-glass viewing screen. L_1 and L_2 are convex lenses of focal length d, and each of the distances TL_1, L_1P, PL_2 and L_2S is equal to d. When P is empty an inverted image of T is thrown on to S. A photographic plate at P records the intensity of the light at each point in its plane, and is developed and printed as a hologram whose transparency at every point is proportional to this intensity. With the hologram in position the inverted image at S is slightly but not seriously blurred; the interesting point is that part of the transparency can now be screened from the laser and the whole of its image is still visible at S. Why?

The basic physical fact we need is that when coherent light is passing through both focal planes of a lens the electric fields in the two planes are Fourier transforms of one another. So if $F(x, y)$ is the field at the transparency, and $f(u, v)$ the field at P, then

$$f(u, v) = \iint F(x, y) \exp 2\pi i(ux+vy)\, dx\, dy.$$

The light intensity at P is the square modulus of f, namely f^*f, where the asterisk denotes the complex conjugate; so the transparency of the hologram is proportional to f^*f, which is just the two-dimensional power spectrum of the pattern $F(x, y)$. Now suppose that F is in two parts, F_1 and F_2, and that we illuminate F_1 but cover up F_2. The wave reaching the hologram will be f_1, the Fourier transform of F_1, and the wave emerging from the other side will be $f_1(f_1^*+f_2^*)(f_1+f_2)$. Of particular concern to us is the component $f_1f_1^*f_2$. When Fourier transformed by the second lens (see Appendix) this gives F_2 (upside down) slightly blurred by convolution with the Fourier transform of $f_1^*f_1$, which is the autocorrelation function of F_1. The reason why the blurring is generally slight is that if F_1 is an effectively random pattern its autocorrelation function will be small and random except in the immediate neighbourhood of the origin, that is, the pattern F_1 will not be a good match with a displaced version of itself. Hence the appearance of F_2 on the viewing screen even when that part of the transparency is screened from illumination by the laser.

Instead of raising the embarrassing question whether the skull can contain anything corresponding to a laser, perhaps we should say something about a mathematically similar model designed to do in the

dimension of time what the holograph does in space. At present perhaps the model is of more logical than neurological significance, but it is too early as yet to be certain.

3. Temporal holography

In Fourier holography we record the power spectrum of a spatial pattern, and use it to modulate the Fourier transform of part of the pattern, recovering the other part by what may be described as Fourier synthesis. Longuet-Higgins (1968*a*, *b*) showed that precisely similar operations could be applied to time-dependent signals, using a device which he has named the 'holophone'. This is essentially a bank of narrow-pass filters, connected in parallel to the input channel, and similarly connected to the output channel through amplifiers of variable gain. The battery of amplifiers corresponds mathematically to the light-sensitive grains of the holographic plate. When a signal is put into the device, each filter transmits a certain amount of energy, and the gain of its amplifier is turned up by a proportional amount. The gains of the amplifiers thus represent the power spectrum of the recorded signal. Several signals can be recorded in this way, and the amplifier gains will then represent the sum of their power spectra. Now suppose that a recorded signal is in two parts, an earlier part $F_1(t)$ and a later part $F_2(t)$. Then arguments very similar to those of the preceding section show that if $F_1(t)$ alone is sent along the input channel, the output channel will emit both $F_1(t)$ and $F_2(t)$, in the correct temporal relation. The later part, F_2 (and for that matter the cue F_1) will be accompanied by a certain amount of noise in the form of an 'echo' unless F_1 happens to be a single sharp pulse, and this echo will be more noisy the shorter the cue and the greater the combined length of the recorded signals (Willshaw & Longuet-Higgins, 1969).

As an information storage and retrieval device the holophone has advantages and disadvantages which correspond exactly with those of the holograph. It can be used for storing several complex signals, though the signal-to-noise ratio falls in proportion to the number stored. The storage is non-local: damage or ablation affects all the stored signals a little, rather than any one of them in particular, or any part of one. And both the holograph and the holophone are content-addressable, in the sense that mere presentation of a 'CS' is enough to evoke the appropriate response, without any need to locate its address in the

system. But both systems suffer from a certain rigidity: a recorded pattern or signal will evoke its partner or proper successor only if it is presented at exactly the right scale (for the holograph) or exactly the right tempo (for the holophone). If one is thinking of the former device in connexion with visual learning, or the latter in connexion with the learning of sequential routines this is obviously a serious limitation. Also, a holophone has to meet rather stringent specifications in order to function properly: it can record a signal of duration D only if (a) the inverse bandwidth of each filter exceeds D, (b) the inverse frequency separation between neighbouring filters also exceeds D, and (c) the resonant frequencies of the filters are not subject to drift. Nevertheless, the possibility of storing temporal signals in frequency space does raise the question whether some parts of the cortex may not code temporal signals in this sort of way, and whether there may not be systems of cells or neural circuits which respond selectively to particular frequencies —most likely in the 100-cycle to 10-cycle range.

4. Correlation models

As already remarked, the output of a holograph, when it is used to store pairs of patterns and to recall a particular one from its partner, is mathematically expressible in terms of convolutions and correlations involving the patterns concerned—and, indeed, those not concerned (see Appendix). It does this by storing the superimposed power spectra of the various pairs, rather than their internal correlations. But there is a well-known theorem to the effect that the power spectrum and the autocorrelation function of a pattern are Fourier transforms of each other, so that to record the one is equivalent, in terms of information, to storing the other. Could one therefore not imagine storing directly the correlations between or within the various input signals, without the need for Fourier-analysing the signals and all the apparatus which that entails? Various authors have pointed out that the performance of the holograph can be crudely mimicked by simple optical devices working with ordinary incoherent light. But the first person to put forward a detailed model of associative memory based on the correlation principle seems to have been Roy (1960, 1962). Roy's papers are concerned with associative recall in time, and describe a device incorporating a large number of delay lines, each of which is involved in recording the autocorrelation for a particular time interval—though Roy

does not say so explicitly. The time scale is quantized, and the information is stored in a number of units rather like potentiometers, but the input-output characteristics of the system are virtually identical with those of the holophone (a later proposal). The same problem of tempo arises—if it really is a problem—and Roy has considered how it might be overcome by cunning neural circuitry (private communication). But although the general idea is attractive, we are still without any quantitative estimate of the storage efficiency of the device, and it is not easy to see how its relevance to the brain could be established by physiological experiments—a dilemma which should not, of course, be regarded as nullifying a good idea.

So, many of the features of holographic recording might be retained, and some of the technical difficulties avoided (Gabor, 1968*a*, *b*), if one could store correlations directly, rather than in the form of power spectra. This theoretical suggestion was explored by Willshaw, Buneman & Longuet-Higgins (1969); and as their analysis led to some useful conclusions about information storage efficiency we shall summarize it as briefly as possible.

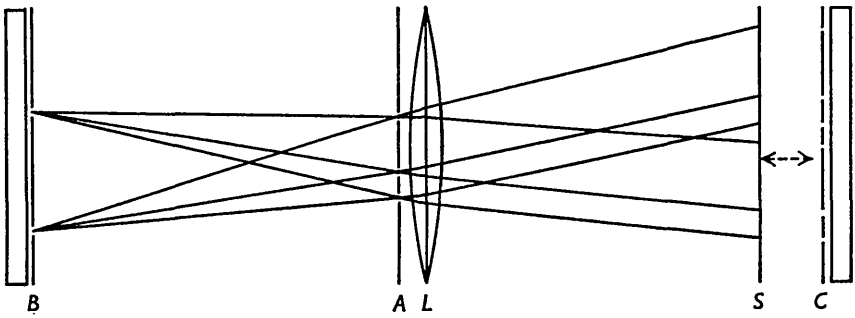

Fig. 2. The distances AB and LS are both equal to the focal length of L. In the reconstruction of B from A, S is replaced by C, which is illuminated from the right.

They began by considering an optical device, the correlograph (see Fig. 2), which could form on a viewing screen S the cross-correlation between two pinhole patterns B and A. In the case illustrated B consists of two pinholes and A of three, so that the 'correlogram' comprises 6 bright spots. The correlogram is copied as a third pinhole pattern C, and is substituted for S, which in turn is mounted where B was before. Light from a diffuse source is now allowed to shine through C and A, and a pattern of bright spots appears on S. But the 18 rays now passing

through do not produce as many bright spots on S; six of them converge in sets of three on to the sites of the original pinholes in B, and the other 12 rays strike S at random points. Using a detector with a threshold of three brightness units we can therefore reconstruct the original B pattern with full accuracy.

They went on to consider the logic of the device. Because of the focusing action of the lens any ray through A and B parallel to a given line will illuminate the same point on S, so that there is a many-to-one mapping from point pairs in A and B on to points in S. This makes it possible to recover a displaced B-pattern from a correspondingly displaced A-pattern, but seriously limits the number of pattern pairs that can be associated in one correlogram. They therefore turned their attention to a logically related but very different-looking model, in which every pair of 'points' in A and B is mapped on to just one point of a third set S. A natural realization of this model is the 'associative net', in which (see Fig. 3) the 'points' of A are input fibres running from left to right, the 'points' of B are output fibres running upwards and those of S are on–off switches at the intersections. The theory is as follows. Let there be N_A A-lines, N_B B-lines and $N_A N_B$ switches. Let a typical A-pattern consist of impulses coming along M_A of the A-lines chosen at random, and let a typical B-pattern likewise involve M_B B-lines. Then an A-impulse will cross a B-impulse at each of $M_A M_B$ intersections, and at each of these the switch is turned on if it is not already on. After a number of pairs of patterns, say R pairs, have been associated in this way, a certain fraction p of the switches will have been turned on; if the patterns are random p will approximately be given by

$$p = 1 - \exp(-RM_A M_B/N_A N_B).$$

After the storage, the recall. An A-pattern is put in, and each B-line of the associated B-pattern receives impulses through M_A switches (which, by hypothesis, have all been turned on). But a B-line not belonging to the B-pattern will also receive a certain number of impulses. The chance that it receives an impulse through every one of its M_A intersections with the active A-lines is in fact p^{M_A}. So if each B-line has a firing threshold equal to M_A, not only will all those lines fire which belong to the B-pattern, but a further $N_B p^{M_A}$ will probably fire as well. The critical value of p will therefore be that for which this number is approximately unity; that is,

$$\log_e N_B = -M_A \log_e p.$$

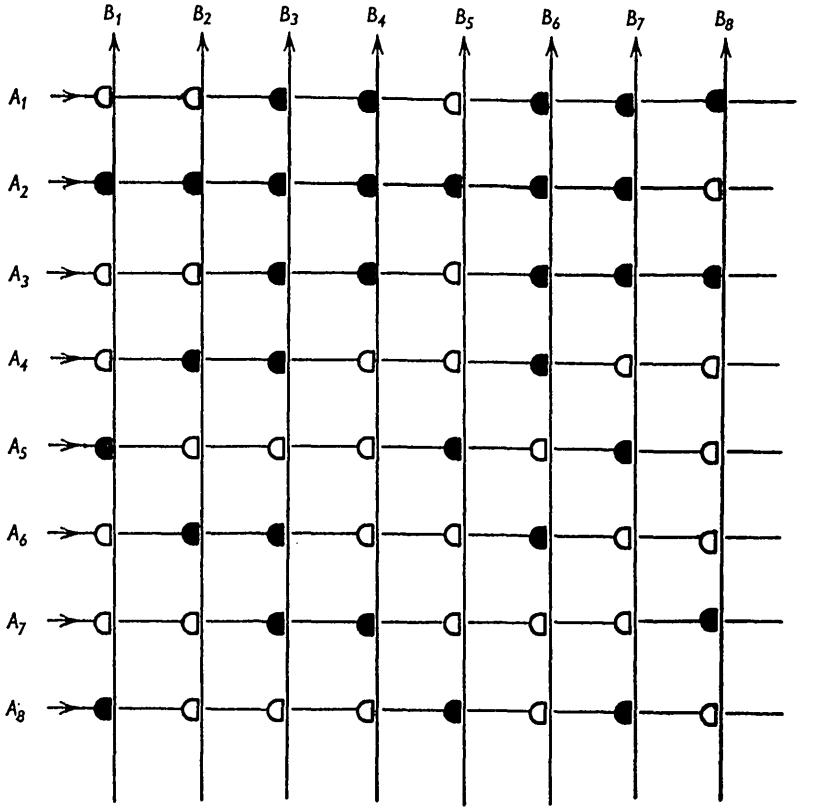

Fig. 3. An associative net. The input lines ($A_1 \ldots A_8$) run horizontally, the output lines ($B_1 \ldots B_8$) run vertically. The filled or open semicircles represent switches which have or have not been turned on. Four associations have been recorded, namely:

A-pattern	B-pattern	A-pattern	B-pattern
1, 2, 3	4, 6, 7	2, 4, 6	2, 3, 6
2, 5, 8	1, 5, 7	1, 3, 7	3, 4, 8

With this value of p, how much information will have been stored? A single B-pattern requires $\log_2 \binom{M_B}{N_B}$ bits to specify it, so if R B-patterns can be accurately retrieved, the available information stored is

$$R \log_2 \binom{N_B}{M_B} = RM_B \log_2 N_B \text{ bits approximately.}$$

Combining these three equations we deduce that the number of bits stored is

$$N_A N_B \log_2 p \log_e (1-p),$$

which has its maximum when $p = \frac{1}{2}$, and is then $0{\cdot}693 N_A N_B$ bits. This maximum will in fact never be attained, because the derivation assumes that N_B is virtually infinite; but it does show that under ideal conditions the associative net is more than 50% efficient in the use it makes of its $N_A N_B$ binary switches. Another conclusion of almost equal importance emerges at once from the second equation above. When $p = \frac{1}{2}$, M_A should be about $\log_2 N_B$ (though this is a lower limit, of course), and therefore much smaller than N_B itself. *The input signals should therefore comprise rather small numbers of pulses if the system is to be an efficient information store.*

In a subsequent paper (Willshaw & Longuet-Higgins, 1970) it has been shown that the associative net can be made to function accurately even if the A-patterns—the 'conditioned stimuli'—are defective in a certain fraction of their M_A impulses, and even if the system is partially ablated, in the sense that some of the switches which should be turned on are taken away or turned off at random. To guard against such handicaps the mean number of impulses in an A-pattern must be raised above the minimal value quoted above, and substantially fewer associations than before can be stored if the output is not to be contaminated with spurious pulses. As a result, the information storage density drops sharply, a phenomenon which one might have anticipated from the Winograd–Cowan theorem (1963), which states that if a logical function is to be implemented with unreliable components, then there is a lower limit to the redundancy which must be built into the system if the output is to be accurate.

To sum up this section, we must say that it is indeed possible to mimic many of the most attractive features of holographic memories with devices which store correlations directly rather than as power spectra. In particular, the associative net resembles the holograph in (*a*) operating with fully parallel logic (*b*) storing its information non-locally and (*c*) allowing the storage of many different associations in the same system. It differs from the holograph in (*d*) not requiring any transformation of the input and output signals, or any coherent source of excitation for this purpose (*e*) employing threshold elements to get rid of unwanted noise in the recall and (*f*) using on–off switches rather than grains of variable transparency for storing the individual bits of information which go to make up the recorded associations. In the last three respects it harmonizes distinctly better with our neurophysiological knowledge, particularly if the Hebb synapse is the basic modifiable

element in the animal cortex. We shall have more to say on this last point in later sections.

5. Adalines and perceptrons

A single output line of an associative net, connected to many input lines, through as many modifiable switches, bears a strong superficial resemblance to the 'Adaline' (adaptive linear classifier) invented by Widrow in 1960 (see Widrow, 1964). The Adaline is a pattern classifying device to which patterns are presented as ordered sets of N numbers $(x_1, x_2, \ldots, x_N)$. There are N input lines, and each x_i in every pattern is either 0 or 1. The pattern $(x_1, x_2, \ldots, x_N)$ may be regarded as a point in N-dimensional space; the two classes to which the patterns are to be assigned may be designated as C and C'. There *may* exist a plane P such that all the points of class C and the point $(0, 0, \ldots, 0)$ lie on the same side of P, and all the points of C' lie on the other side of P; if so, the patterns are said to be linearly separable. In this case the Adaline can be taught to separate them—to emit a pulse when presented with a pattern of class C', and not to emit a pulse for a pattern of class C. What it does is to form the sum

$$w_1 x_1 + w_2 x_2 + \ldots + w_N x_N,$$

where the w_i are the current values of N adjustable weights. If this sum exceeds the current value of an adjustable threshold t, a pulse is output; otherwise not. If when a pattern of class C is presented a pulse is (incorrectly) output, the operator lowers w_i by a predetermined amount δw if x_i was 1, and raises t by a predetermined amount δt, and presents the incorrectly classified pattern again. If, on the other hand, a pattern of class C' fails to provoke a pulse, the threshold is lowered and the weights raised, by the same amounts. If a pattern is correctly classified, the threshold and the weights are left as they were. Ultimately, it can be proved, the Adaline responds correctly to every pattern.

An output line of an associative net resembles an Adaline in forming a weighted sum, at any stage, of the 1's and 0's coming along its input lines and in outputting a pulse, or not, according as the resulting sum exceeds or falls short of a certain threshold. And if a pulse ought to be output, those weights which are not already 1 are altered from 0 to 1. But there are important differences, which must not be overlooked.

First, the threshold of an associative net is, by hypothesis, fixed in magnitude—a hypothesis which may of course be invalid for real neural networks. Secondly, each weight in the associative net is assumed to be 0 or 1; if a switch is off $w_i = 0$, and if it is on $w_i = 1$. And thirdly, the switches in the associative net can only be switched on, not off. These handicaps might occasion some surprise that the net can work so efficiently under good conditions. The explanation seems to be that when about half the switches have been turned on, most of them will either be off, or be on because just one association demands it. For this statement to hold, the number of stored associations multiplied by the number of 1's in an input signal must not exceed the number of input lines—a conclusion to which Marr (1969) and Brindley (1969) have been led in recent neurological studies.

It is not possible, however, to teach an Adaline or an output line of an associative net to classify a given set of patterns in an entirely arbitrary fashion. For large N most arbitrary divisions of a large set of patterns into patterns of class C and patterns of class C' are such that no plane can be drawn between the C points and the C' points. The corresponding limitation for the associative net is that if two input patterns both fire a given output line, then any third pattern whose 1's are entirely included among those of the first two patterns, must also fire that line. So to achieve more general classifications some further apparatus is needed. In their book on perceptrons* Minsky & Papert (1969) deal elegantly with this general problem; but we must be content to refer to the fundamental theorem on which their work is based. To state the theorem we need the idea of a 'mask'. A mask is a device which can be applied to a pattern and outputs a 1 if *all* the x_i belonging to a given subset are equal to 1; otherwise it outputs a 0. Then the theorem states that, given any classification of a set of input patterns into patterns of class C and patterns of class C', there exists a set of masks such that if each mask is connected to one input line of an Adaline, whose weights and thresholds are suitably adjusted, the system will output a pulse if and only if the input pattern is of class C'. Such a two-layer system, incorporating a set of masks and a linear threshold device, is called a two-layer perceptron. Informally speaking, the first layer is the 'coder', and the second layer, the Adaline, is the 'memory', since it is there that the learning, if any, takes place. The problem of choosing a suitable set of

* The original perceptron was invented by Rosenblatt (1958, 1962), but since that time the word has widened considerably in application.

masks for a particular classification task is a problem of great interest, and is Minsky and Papert's main concern in their book. For us the essential point is that a two-layer perceptron *can* learn any classification task if it is equipped with an appropriate set of masks; and we shall see that the concept of a mask in a two-layer perceptron is virtually identical with Marr's concept of a 'codon' in his theory of the cerebellum (1969), to which we now turn.

6. The cerebellar cortex

The cerebellar cortex is one of the most tantalizing organs of the brain, because of its extremely orderly structure, which has been the subject of intensive study ever since Ramón y Cajal (1911) first mapped its main features. In bald and inadequate outline, the cerebellum has (Eccles, Ito & Szentegothai, 1967) one set of outputs—the inhibitory outputs of the Purkinje cells—and two sets of inputs. The simpler set of inputs is along the climbing fibres, each of which is polysynaptically connected to just one Purkinje cell, and is capable of exciting the Purkinje cell apparently without assistance. The more complex set of inputs is along a much greater number of mossy fibres, which are connected in a few-to-few fashion to the dendrites of the extremely numerous granule cells. The axons of the granule cells are monosynaptically connected to the dendrites or the bodies of the Purkinje cells. These axons are the parallel fibres; and whereas each parallel fibre may form a synapse with perhaps 500 Purkinje cells, each Purkinje cell receives synaptic connexions from literally hundreds of thousands of parallel fibres. (Also ministering in some way to the needs of the Purkinje cells are the stellate cells and the basket cells, which are also controlled by the parallel fibres; and the activity in the parallel fibres is regulated by Golgi cells, which sample the mossy fibre activity and also that of the granule cells. We may ignore these complications for the moment.)

The crux of Marr's theory (1969) is that the Purkinje cells are the output lines of an associative net (though his paper slightly predates that of Willshaw *et al.* referred to in §4). A given Purkinje cell can be made to fire by the appropriate UCS, provided by its climbing fibre; but if this firing is accompanied by a CS in the form of a parallel fibre input, then the CS alone will later suffice to fire the Purkinje cell. This process of conditioning is achieved, in Marr's view, by facilitation of the synapses

between the relevant parallel fibres and the Purkinje cell; if the Purkinje cell threshold is low enough a set of impulses along these fibres will fire it without any input from the climbing fibre. Marr therefore postulates that the parallel fibre-Purkinje cell synapses are potentially excitatory and can be activated in the manner of Hebb synapses—a nice firm experimental prediction.

Marr's second hypothesis concerns the translation from mossy fibre input into parallel fibre activity. This process he regards as fulfilling the same function as the masks in the two-layer perceptron (see §5), though he uses the phrase 'codon representation' to designate the transformation to which a particular mossy fibre input is subjected before being referred to the Purkinje cells. He gives quantitative reasons why a granule cell should receive rather few synapses from mossy fibres—why, in other words, patterns can be rather well separated with quite few masks.

According to this view, which harmonizes most attractively with the known anatomy, the cerebellar cortex behaves as a battery of two-layer perceptrons, if the word 'perceptron' is allowed to encompass a learning machine in which a particular weight, once altered, cannot be altered back again. The system is supposed to learn two kinds of task, under cerebral instruction: the performance of complex movements, and the maintenance of posture and balance. The primary instructions (UCS) are received along the climbing fibres, which originate in the cells of the inferior olivary nucleus; whenever an olivary cell fires, it sends an impulse to its Purkinje cell. The Purkinje cell is also provided, via its mossy fibre input, with information (CS) about the context in which its olivary cell fired. Later, when the action has been learnt, occurrence of the context alone is enough to fire the Purkinje cell, which then initiates the next elementary movement; the action thus progresses as it did during rehearsal.

The suggestion that the cerebellum learns motor skills in this way is due to Brindley (1964); but Marr develops the idea in considerable detail, both mathematical and anatomical. In particular, he works out the possible range of codon sizes, and finds that they should be very small (4 or 5) in relation to the number of active mossy fibres (500–1500). He also points out that the stellate and basket cells, which inhibit the Purkinje cells, may very well serve the purpose of adjusting the Purkinje cell threshold—downwards if the mossy fibre or granule cell activity is unusually high; and that the Golgi cells may perhaps act similarly

in controlling the codon size. (These hypotheses, though appealing, are not crucial to the theory.) He does not, however, consider the Purkinje cells collectively, and therefore draws no conclusion about the density, in bits per modifiable synapse, with which the available information is stored. The amount of available information, in the absence of redundancy, is $R \log_2 \binom{N}{M}$, where N is the number of output lines, M the number which were active in any response and R the number of learned responses. But the analysis of §4 showed that for an associative net with constant output thresholds this amount of information was indeed comparable with the total number of adjustable switches; so there can be little doubt that Marr's model of the cerebellar cortex is capable of storing its output information efficiently—without wasting too many modifiable synapses.

7. Sequential models

Similar ideas to those underlying Marr's theory have been used by Brindley (1969) in the construction of nerve-net models which will learn large numbers of elementary 'word' sequences, and are economical in storage space and realistic in their neurological components. The task is to memorize say 10^5 three-word sequences composed from a vocabulary of say 10^4 words, so that when the first two words of a sequence are supplied the third will be reliably recalled. The proposed models store their information in modifiable synapses; their connexions need to be specified only in a general way, and they can tolerate the destruction of many cells. The model which Brindley favours most is built with Hebb synapses, and demands an observationally plausible number of inputs to each cell. Brindley shows that this model stores its information economically—that it does not waste a large fraction of its modifiable synapses—and suggests that such economy can only be achieved if there is an abundance of cells with many independent inputs and rather low thresholds. This conclusion harmonizes very satisfactorily with the theory of the associative net. Brindley deals with the problem of timing by postulating delay lines and repeater cells which keep the first 'word' in hand till the second word arrives. Doubtless there are such components in our temporal memories, but there is obviously room for much more theoretical work on timing devices, particularly in relation to the question of variability of tempo (see §§3 and 4). A final

point in favour of Brindley's scheme is its embryological plausibility, for which he presents detailed arguments; but these, regrettably, would take us outside the scope of this review.

8. Discussion

In this article we have not attempted to review, or even to refer to, all the good work which has been devoted to the study of the memory. In particular we have not felt competent to undertake any discussion of the abundant psychological evidence relating to associative memory, or to place the various theoretical models in any precise relation to the anatomy of the brain, except in so far as this has already been done by Marr for the cerebellar cortex. We have concentrated on one particular problem: how is it possible to construct, with mainly parallel logic, an associative information store which can learn to associate very many pairs of conditioned and unconditioned stimuli, when the individual stimuli are so complex that they cannot be anticipated in any detail before they arrive? And how can this task be performed with the maximum economy in the use of modifiable elements? In reviewing the published work on this problem we have been struck by the way in which people in quite diverse fields have converged upon the same general ideas: namely that this end is well served by storing the correlations between the separate components of each CS and the corresponding UCS; that it is entirely satisfactory to store these correlations indiscriminately in a large number of binary stores; that some kind of threshold logic is essential if one is to obtain accurate responses after learning; and that systems of this general design can be made to store the relevant information at a high density in bits per register. Furthermore, such systems can be made tolerant to extensive damage, or to inaccuracy in the conditioned stimuli, by a corresponding sacrifice in information storage density—a possibility which arises from the distributed manner in which each individual association is stored. And last but not least, it appears that the cerebellum may very well be an information-processing system of just this kind.

Finally, a word about the particular problem of temporal recall. Just as the holograph can be mimicked by the correlograph and the associative net, so the holophone, which stores temporal variations in frequency space, can be mimicked by a system incorporating a large number of delay lines. We really have insufficient evidence at present on

which to base a theoretical decision; frequency analysis is certainly a neurological possibility, as something of the kind is obviously done by the ear, but the finite speeds of neural impulses would provide an obvious basis for neural models based on the delay principle. Perhaps both types of coding are used in the brain; experiment alone can decide.

MATHEMATICAL APPENDIX

Fourier holography

As remarked in §4, the electric field amplitude $f(u, v)$ reaching the plane P is the Fourier transform of that leaving T. That is,

$$f(u, v) = \iint F(x, y) \exp\{2\pi i(ux+vy)\}\, dx\, dy,$$

where u and v are holographic coordinates, which have been normalized by dividing by the laser wavelength and by the focal length of L_1. A second Fourier transformation is effected by L_2:

$$\begin{aligned} F'(x, y) &= \iint f(u, v) \exp\{2\pi i(xu+yv)\}\, du\, dv \\ &= F(-x, -y), \end{aligned}$$

and gives the inverted image F' on the screen S in the absence of the hologram. With the hologram in position, illuminated by light passing through the first part of the transparency, the light reaching S is the Fourier transform of $f_1 f^* f$, where $f = f_1 + f_2$. To determine the contribution of the particular term $f_1 f_1^* f_2$ we use the theorem which states that the transform of a product is the convolution of the transforms of the individual factors. It follows that the transform of $f_1 f_1^*$ is

$$\begin{aligned} &\iint F_1'(x-x', y-y')\, F_1(x', y')\, dx'\, dy' \\ &\qquad = \iint F_1(x'-x, y'-y)\, F_1(x', y')\, dx'\, dy', \end{aligned}$$

which is the autocorrelation function of F_1, and may be written as $C(x, y)$. Convolving this with F_2', the Fourier transform of f_2, we obtain

$$\iint C(x-x', y-y')\, F_2'(x', y')\, dx'\, dy',$$

a blurred, inverted image of F_2. If F_1 is irregular, $C(x, y)$ will be small and random except at the origin, and in these circumstances F_2 will be discernible in spite of the blurring, especially if it is made up of a number of small bright spots.

An important question about the holograph, used as an associative

memory, concerns the signal-to-noise ratio of the output signal evoked by a given input. The same problem arises in any device which stores correlations between pairs of patterns; the more pairs that are stored, the noisier the recall. Suppose (to change our notation slightly) that a holograph or linear correlograph has been loaded with R pairs of patterns, F_1 and G_1, F_2 and G_2, etc. Then the input of F_1 will evoke the composite output

$$F_1 \% (H_1' \% H_1 + H_2' \% H_2 + \ldots),$$

where the % sign has been used for convolution, a dash denotes inversion of a pattern, and

$$H_1 = F_1 + G_1, \quad H_2 = F_2 + G_2, \quad \text{etc.}$$

When this expression is expanded, just one of the $4R$ terms is that which enables F_1 to recall G_1, namely $F_1 \% F_1' \% G_1$; all the others count as noise. More detailed analysis shows that the signal-to-noise ratio in the recall of G_1, if all the patterns are chosen completely at random, and are of comparable intensity, is the area covered by F_1 divided by the combined area of all the recorded patterns. This fact underlines the necessity of filtering the output through a battery of threshold elements, such elements being not only practically but logically necessary if the system is to be used for *recognizing* patterns rather than just recalling them.

References

Beurle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses. *Phil. Trans. R. Soc. Lond.* B **240**, 55.

Brindley, G. S. (1964). The use made by the cerebellum of the information that it receives from sense organs. *Int. Brain. Res. Org. Bull.* **3**, 80.

Brindley, G. S. (1969). Nerve net models of plausible size that perform many simple learning tasks. *Proc. R. Soc. Lond.* B **174**, 173.

Cragg, E. C. & Temperley, H. N. V. (1954). The organisation of neurones; a co-operative analogy. *Electroenceph. clin. Neurophysiol.* **6**, 85.

Eccles, J. C., Ito, M. & Szentagothai, J. (1967). *The Cerebellum as a Neuronal Machine.* Berlin: Springer-Verlag.

Gabor, D. (1948). A new microscopic principle. *Nature. Lond.* **161**, 777.

Gabor, D. (1949). Microscopy by reconstructed wavefronts. *Proc. R. Soc. Lond.* A **197**, 187.

Gabor, D. (1951). Microscopy by reconstructed wavefronts. II. *Proc. phys. Soc.* **64**, 449.

Gabor, D. (1968). Improved holographic model of temporal recall. *Nature, Lond.* **217**, 1288.

GABOR, D. (1969). Associative holographic memories. *IBM. Jl Res. Dev.* **13**, 156.

GRIFFITH, J. S. (1963). A field theory of neural nets. I. Derivation of field equations. *Bull. math. Biophys.* **25**, 111.

GRIFFITH, J. S. (1965). A field theory of neural nets. II. Properties of the field equations. *Bull. math. Biophys.* **27**, 187.

HEBB, D. O. (1949). *The Organisation of Behaviour.* New York: Wiley.

VAN HEERDEN, P. J. (1963*a*). A new optical method of storing and retrieving information. *Appl. Optics.* **2**, 387.

VAN HEERDEN, P. J. (1963*b*). Theory of optical information storage in solids. *Appl. Optics* **2**, 393.

LEITH, E. N. & UPATNIEKS, J. (1962). Reconstructed wavefronts and communication theory. *J. opt. Soc. Am.* **52**, 1123.

LONGUET-HIGGINS, H. C. (1968*a*). Holographic model of temporal recall. *Nature, Lond.* **217**, 104.

LONGUET-HIGGINS, H. C. (1968*b*). The non-local storage of temporal information. *Proc. R. Soc. Lond.* B **171**, 327.

MARR, D. (1969). A theory of cerebellar cortex. *J. Physiol., Lond.* **202**, 437.

MINSKY, M. & PAPERT, S. (1969). *Perceptrons.* Boston: M.I.T. Press.

PRIBRAM, K. H. (1966). Some dimensions of remembering: steps towards a neuropsychological model of memory. In *Macromolecules and Behaviour* (ed. J. Gaito). New York: Appleton–Century–Crofts.

PRIBRAM, K. H. (1969). The neurophysiology of remembering. *Scient. Am.* **220**, 73.

RAMÓN Y CAJAL (1911). *Histologie du Système Nerveux de l'Homme et des Vertébrés.* Paris: A. Maloine.

ROSENBLATT, F. (1958). The perceptron. A probabilistic model for information storage and organisation in the brain. *Psychol. Rev.* **65**, 386.

ROSENBLATT, F. (1962). *Principles of Neurodynamics.* Washington D.C.: Spartan Books.

ROY, A. E. (1960). On a method of storing information. *Bull. math. Biophys.* **22**, 139.

ROY, A. E. (1962). On a method of storing information. II. A further study of model properties. *Bull. math. Biophys.* **24**, 39.

STROKE, G. W. (1966). *An Introduction to Coherent Optics and Holography.* New York: Academic Press.

WESTLAKE, P. R. (1967). Towards a theory of brain functioning: the possibilities of neural holographic processes. *Conferences Proc. 20th Annual Conference on Engineering in Medicine & Biology, I.E.E.E.*

WIDROW, B. (1964). Pattern recognition and adaptive control. *Trans. A.I.E.E.E. Appl. Indust.* **83**, 269.

WILLSHAW, D. J., BUNEMAN, O. P. & LONGUET-HIGGINS, H. C. (1969). Non-holographic associative memory. *Nature, Lond.* **222**, 960.

WILLSHAW, D. J. & LONGUET-HIGGINS, H. C. (1969). The holophone–recent developments. In *Machine Intelligence*, 4 (ed. D. Michie). Edinburgh University Press.

WILLSHAW, D. J. & LONGUET-HIGGINS, H. C. (1970). Associative memory models. In *Machine Intelligence*, 5. Edinburgh University Press (in the Press).

WINOGRAD, S. & COWAN, J. D. (1963). *Reliable Computation in the Presence of Noise*. Boston: M.I.T. Press.