An After-Shannon Measure of the Storage Capacity of an Associative Noise-like Coding Memory*

Sergio Bottini

Consiglio Nazionale delle Ricerche
Istituto di Elaborazione della Informazione, Pisa, Italy

Abstract. The maximum amount of information that can be stored, on the average, in each storage element, according to an associative scheme, has been measured for the memory model proposed by the author (Bottini, 1980). In this model, the items being stored are coded by noise-like keys and the memory traces formed in this way are superimposed on the same storage elements. It will be shown that the problem of measuring the information retrieved from the memory in a single recall and the problem - concerning the data-communication field - of measuring the information transmitted over a noisy channel are formally similar. In particular, the Shannon noisy-channel coding theorem can find an application also in our case of an associative memory.

*Submitted to BIOLOGICAL CYBERNETICS - october 1986

1 Introduction

There have been several independent approaches to the mathematical modeling of associative memories (for a review see Kohonen, 1977; Longuet-Higgins et al., 1970). The models produced by these approaches are essentially of two types: the so-called matrix models (see Steinbuch, 1961; Kohonen, 1972); and the convolution-correlation models (see Gabor 1968a, b, 1969; Bottini, 1977, 1979, 1980), which originated from the optical holographic models of memory. Despite the independence of their origins, all these models are based on a common principle. In every case, in fact, the success in recalling stored information depends on the degree of correlation between the recall key and the key used, on storage, for coding that information.

These models memorize information in discrete storage elements, each of which has a certain number of levels, i.e. possible states. There can be, for example, two variants of a same memory system, one which requires two-level storage elements, and the other with many-level storage elements. The choice of a particular type of storage element is in strict relation to both the type of item that can be stored and the type of key to be used for storage and recall. The properties of the model will, of course, depend on the nature of the items, keys, and storage elements chosen. This is particularly true for one of the most important attributes of an associative memory, i.e. its storage capacity, which is defined as the maximum amount of information that can be stored, on the average, in each of its storage elements.

In the present paper, the storage capacity of an associative memory endowed with many-level storage elements will be calculated. The model that will be studied in detail is the convolution-correlation model proposed by the author. The results obtained could however be easily transferred to the matrix models. For the storage capacity of a memory system with binary storage elements the reader is referred to the studies made by Palm (1980) and by Willshaw et al. (1969).

To calculate the storage capacity of the model we will, first, measure the amount of information retrieved from the memory in a single recall and then sum the contributes of all the memory traces to obtain the total amount of information stored; finally, we will evaluate the maximum value for this stored information.

The method used here to measure the amount of information retrieved in a single recall retraces the techniques introduced by Shannon into the data-communication field to measure the information transmission rate for a noisy channel. As will be shown, in fact, the problems of measuring the relevant amounts of information in the two cases are formally similar.

2 Description of the Memory Model

The model (Bottini, 1980), which presents strong analogies with biological associative memory, was developed from a two-step mathematical transformation suggested by Gabor. Gabor noted that if a given vector f is convolved with a noise-like vector ν, then the inverse operation capable of (nearly) recovering the vector f from the result of the convolution is the correlation with the same noise-like vector ν (or part of it). A noise-like vector is a sufficiently complex, zero mean, not periodically behaved vector. The relevance of these two operations to neural and theoretical information processing systems has also been studied by Borsellino and Poggio (1972,1973).

The central point of the model is a distributed noise-like coding of the information. On storage, each of two associated items codes the other by means of a noise-like key, obtained from the item itself through a randomizing preprocessing. All the memory traces are superimposed on the same memory vector. The noise-like key's property of correlating sharply with itself or with a fragment of itself and, at the same time, not correlating at all with a different key, makes selective retrieval of the information stored possible.

In order to calculate exactly the performance of the model, storage and recall noise keys coming from an ideal random process will be assumed here. Moreover, although in the model the recall process was recycled in order to improve the storage capacity, the restricted case of the one-step-only recall will be studied. Let us then first give the basic formalism of the model.

The information items dealt with by the model are N_0-dimensional vectors, whose elements $I_r(l)$ ($l=1,\ldots,N_0$; $r=1,\ldots,K$, with K the number of the items to be stored) are, for the sake of simplicity, binary, with values 0 or 1. The model could, however, process item elements whose possible values are more than two (real-valued elements could also be permitted). Dimension N_0 will be equal to or less than the number N of the storage elements.

The noise keys used for coding the information items for storage and for recalling them are N-dimensional vectors having as elements, $\nu_r(i)$, zero mean, statistically independent, two-value random variables. These two values are opposite. (However, keys with real-valued elements could also be used.)

The orthonormality condition for the noise keys is given by

$$E\left\{\nu_r(i)\ \ \nu_{r'}(i')\right\} = \delta_{rr'}\delta_{ii'}/N \qquad , \tag{1}$$

with $i,i' = 1,\ldots,N$, and $r,r' = 1,\ldots,K$; where E denotes the symbol of expectation, ν_r and $\nu_{r'}$ are uncorrelated vectors if $r \neq r'$, and δ is the Kronecker symbol which has the value of 1 for an equal value of the two indices, otherwise it has the value of 0. If a key $\tilde{\nu}_r$ is obtained from a given key ν_r by setting $(1-g)N$

of its elements to zero randomly, with $0 < g \leqslant 1$, and replacing the cancelled elements with extraneous noise, then we have

$$E \left\{ \tilde{\nu}_r(i) \ \nu_r(i) \right\} = g/N \quad . \tag{2}$$

The memory trace corresponding to item I_r is an N-dimensional vector built by the circulant convolution of I_r itself with the associated key ν_r

$$m_r(j) = \sum_{l=1}^{N_0} \nu_r(j-l) \ I_r(l) \quad , \tag{3}$$

where $j = 1,\ldots,N$.

The total memory vector M is obtained from the single K traces through an element by element addition

$$M(j) = \sum_{r=1}^{K} m_r(j) \quad . \tag{4}$$

The values taken on by the elements of M, i.e. our storage elements, are scattered around zero over a number of discrete levels separated by $1/N^{1/2}$.

The recall by means of the noise key $\tilde{\nu}_\varrho$ is given by the circulant correlation of it with the memory M

$$R_\varrho(h) = \sum_{i=1}^{N} \tilde{\nu}_\varrho(i-h) \ M(i) \quad , \tag{5}$$

where $h = 1,\ldots,N_0$. If $\tilde{\nu}_\varrho$ coincides exactly or partially, according to weight g, with the noise key used in one of the storages, then the expected value of the recall will be

$$E \left\{ R_\varrho(h) \right\} = g \ I_\varrho(h) \quad , \tag{6}$$

i.e., apart from the attenuating factor g, the very item coded for storage by that key. Otherwise, if the recall key is totally uncorrelated with respect to all the storage keys, the expected value of the recall will be identically zero.

Since recall is affected by unwanted noise the retrieved values will fluctuate (correspondingly) around 0 or g. A threshold detection operation working at a suitable value Th, with

$$0 < Th < g \qquad , \qquad (7)$$

is then required to restore the original values 0 or 1.

3 Storage Capacity of the Memory

3.1 Recall Error Probabilities

In order to answer the question "How much information can be stored in this memory system?" the probabilities of the two types of error which can occur when recalling an item, namely a 0 retrieved as 1 or a 1 retrieved as 0, must first be known. These probabilities can be determined starting from the variance of recall R. For this variance we have

$$\sigma^2 \simeq \mathcal{I}/N \qquad , \qquad (8)$$

with a good approximation in all cases of interest, i.e. $\mathcal{I} >> 1$ and K not small; where $\mathcal{I}$ equals the sum of the square values of the elements of all the items stored, and N is the number of storage elements. For our binary items, $\mathcal{I}$ simply coincides with the total number of the 1's of the items stored (for the exact calculation of the recall variance see Bottini and March, 1985).

The smaller the σ^2, the higher the probability that the threshold operation can successfully separate the 1's from the 0's for the item recalled. The distribution of the system output (recall) is assumed to be normal because of the large number of terms contributing to it. We can thus determine

$$\mathscr{P}_1 = 1/2 - \Phi(Th/\sigma) \quad , \tag{9}$$

the probability (< 0.5) that for any binary item stored one given 0 can be wrongly retrieved as 1, and

$$\mathscr{P}_2 = 1/2 - \Phi([g-Th]/\sigma) \quad , \tag{10}$$

the probability (< 0.5) that one given 1 can become wrongly 0, with Φ the error function defined as

$$\Phi(x) = (2\pi)^{-1/2} \int_0^x \exp(-y^2/2)\, dy \quad . \tag{11}$$

3.2 The Information Retrieved in a Single Recall

By means of $\mathscr{P}_1$ and $\mathscr{P}_2$ we can now measure the average information H_m retrieved from the memory in a single recall, i.e. the amount of information retrieved per element of R_ϱ on the average.

Let us first calculate the average information H required to describe the process that produces output R_ϱ from the corresponding item I_ϱ , through the storage and recall steps. In this process there are four possible events: i) a 0 is correctly retrieved as 0; ii) a 0 is wrongly retrieved as 1; iii) a 1 is correctly retrieved as 1; and iv) a 1 is wrongly retrieved as 0. Let μ be the density of the 1's – or the probability of having a 1 – for the item considered. The probabilities of occurrence for these four events are then respectively: i) $p_1 = (1 - \mathscr{P}_1)(1-\mu)$; ii) $p_2 = \mathscr{P}_1(1-\mu)$; iii) $p_3 = (1-\mathscr{P}_2)\mu$; and iv) $p_4 = \mathscr{P}_2\mu$. H is obtained from these four probabilities by means of the Shannon formula (with N_0 sufficiently large)

$$H = - \sum_{i=1}^{4} p_i \log_2 p_i \qquad (12)$$

We then measure the average information contained in the item (H_I) and the average information available at the output (H_R). H_I is obtained from the probabilities μ and $1-\mu$; and H_R from $p_R = (1-\mathscr{P}_2)\mu + \mathscr{P}_1(1-\mu)$ and $1-p_R$.

Then

$$H_m = H_I - (H - H_R) \qquad (13)$$

gives the average information, expressed in bits per element of R_ϱ, actually retrieved in a given recall. The accuracy of a given recall can thus be evaluated by means of a parameter, the recall efficiency η_R, which is defined as follows

$$\eta_R = H_m/H_I \qquad (14)$$

This parameter equals the percentage of the item information which is actually retrieved.

Equation (13) has an analog in the communications-system theory. Indeed, as known from Shannon (1948), if $\mathscr{P}_1$ and $\mathscr{P}_2$ were the probabilities with which an error can occur during the transmission of the two symbol types of a message over a noisy channel, and $1-\mu$ and μ were the respective probabilities of production of these two symbols by the source, then H_m in (13) would correspond to the amount of information transmitted per symbol on the average, H_I being the information of the source message and $H-H_R$ the information lost because of noise.

3.3 The Total Information Stored in the Memory

H_m is an amount of information averaged over all the elements of R_ϱ. Thus, if I_ϱ and (accordingly) R_ϱ have $N_0 < N$ elements, with N the number of storage elements, then the average information H_m^* stored in the corresponding memory trace will be given by

$$H_m^* = \lambda H_m \quad , \tag{15}$$

where $\lambda = N_0/N$.

The total average information H_{TOT} stored in the memory vector M is then obtained as

$$H_{TOT} = \sum_{r=1}^{K} H_m^*(\mu_r) \quad , \tag{16}$$

which is expressed in bits per storage element, where K is the total number of traces, and μ_r is the density of the 1's for item I_r .

Let μ_r have the same value μ for any r so that every item has the same number of 1's. It is found then that H_{TOT} converges to the following limit as K (and accordingly N) approaches infinity, or equivalently μ approaches zero,

$$H_{TOT}^{\infty} = (\mathcal{I}/N)\left\{ \log_2\left[(1-\mathcal{P}_2)/\mathcal{P}_1\right] + \mathcal{P}_2 \log_2\left[\mathcal{P}_1\mathcal{P}_2/\left[(1-\mathcal{P}_1)(1-\mathcal{P}_2)\right]\right]\right\} . \tag{17}$$

This value is the upper bound of H_{TOT} for a given choice of the ratio $\mathcal{I}/N$,and of g and Th.

Provided that the ratio $\mathcal{I}/N$, and g, Th, and K remain the same, H_{TOT} increases as λ increases, while its limit value given by (17) (i.e. when K is very large) does not depend on λ.

3.4 Maximum Values for the Information Stored

Finally, another limit value for H_{TOT}, which is useful to complete the description of the quantitative behaviour of the model, is obtained by letting the ratio $\mathcal{J}/N$ approach infinity, while keeping μ fixed (so that K must also approach infinity). For Th = 0.5g , with g $(0 < g \leqslant 1)$ the previously defined weight describing the degree of completeness of the key, we have

$$^{\infty}H_{TOT} = g^2 \left[(\log_2 e)/\pi\right](1-\mu) \qquad . \qquad (18)$$

When $\mu = 0.5$, (18) states that no more than $g^2 0.23$ bits per storage element on the average can be stored. This limit value is therefore the storage capacity of the model when the items have as many 1's as 0's. When μ approaches zero, (18) yields twice as much as this value. However, in such a case, because $\mu < 0.5$, an optimal choice of Th ($> 0.5g$) can be made to maximize H_{TOT} and thus improve its limit value with respect to the value given by (18) for Th = 0.5g. From (17), for a suitable choice of Th, we find that this improvement is, at most, of the order of ten percent. Thus, approximately $g^2 0.5$ bit/element can, at most, be stored in this memory system. This value, achievable only when the items have very few 1's, is the upper bound for the storage capacity. For intermediate values of μ, the storage capacity can be obtained, with a very good approximation, by interpolating linearly the two values found for the extreme cases just considered.

4 Number of Levels Required by Each Storage Element

4.1 Maximum Number of Levels

Now that we have calculated the number of bits per element that can be stored associatively in the memory vector M, it is of interest to calculate how many bits are required by each of these elements to allow the successive storages. As a consequence of the use of the noise keys defined in Sect. 2, the storage of a single item may cause modifications in the values of all the storage elements. This is, therefore, the least favourable condition to save levels (or bits) in the storage elements.

It is easy to see that, if for each storage element we use a number of levels (separated, as already mentioned, by $1/N^{1/2}$, i.e. the absolute value of the noise key elements) given for example by $6\mathcal{I}^{1/2}$, which is required to cover the interval $\left[-3(\mathcal{I}/N)^{1/2}, +3(\mathcal{I}/N)^{1/2}\right]$, then H_{TOT} will be reduced negligibly. In fact, this interval equals six times the standard deviation of the distribution (with variance $\mathcal{I}/N$) of the values assumed by the storage elements, so that the saturation of a storage element occurs only for a negligible fraction of these elements, given by $2\left[1/2 - \Phi(3)\right] = 0.003$, with Φ the error function. Saturation means that a storage element is forced to retain its state because the level being requested is outside the range of the levels permitted. It is thus sufficient that each storage element has a number n of bits given by the smallest integer satisfying

$$n \geqslant \log_2(6\mathcal{I}^{1/2}) \quad . \qquad (19)$$

4.2 Saving Levels by Quantizing the Memory Traces

A smaller number of bits for each storage element can be used if every trace is quantized into two levels - plus the zero level if necessary - before storage.

Let us still consider items having the same number $\mathcal{L}$ ($>>1$) of 1's so that the variance of every memory trace will have the same value σ_{tr}^2. Let us then transform each memory trace vector m_r into a quantized vector $\hat{m}_r$ as follows

$$\hat{m}_r(i) = \varepsilon \operatorname{sgn}\left[m_r(i)\right] \quad , \qquad (20)$$

where sgn is the function that takes the sign of the argument. The value to be given to ε is

$$\varepsilon \simeq (\pi/2)^{1/2} \sigma_{tr} \quad . \qquad (21)$$

for the recall to be correctly normalized (Bottini and March, 1985). The memory traces so obtained are then added together to give the memory vector $\hat{M}$. The variance of the single memory trace $\hat{m}_r$ is ε^2. Therefore, as the corresponding elements of the various memory traces are statistically independent, the variance of $\hat{M}$ is K times ε^2, with K the number of traces. The number $\hat{n}$ of bits for each element of $\hat{M}$ which suffices in this case, i.e. when the values of these elements are quantized into levels which are multiples of ε, is then given by the smallest integer satisfying

$$\hat{n} \geqslant \log_2(6K^{1/2}) \quad , \qquad (22)$$

with $6K^{1/2}$ the number of levels. $\hat{n}$ is thus related to the number K of items stored, irrespective of the value of $\mathscr{L}$. Less levels than before are needed because here the storage of a single item can, at most, cause a storage element to change its state from a given level to an adjacent one, while, without quantization, jumps involving non adjacent levels are possible.

However, the advantage of providing each storage element with a smaller number of bits is here paid for by a reduction in the amount of information stored in the memory. As a matter of fact, within the limits of approximation of (8), the recall variance is now given by

$$\hat{\sigma}^2 \simeq (\pi/2) \mathscr{I}/N \quad , \qquad (23)$$

The increase in the variance value by the factor $\pi/2$ with respect to the value before quantization implies higher values for the recall error probabilities $\hat{\mathscr{P}}_1$ and $\hat{\mathscr{P}}_2$, thus a smaller amount of information can be stored. The right values for information stored are still obtained from (16) and (17), provided that these two new values for the error probabilities are used. In particular, the limit value given by (18) must be here multiplied by $2/\pi$.

4.3 A Two-level Read-only Memory

On the other hand, instead of quantizing singly every memory trace before storing it, we can quantize the total memory into two levels, in just one step, after the memory has been formed as normal. The values to be assigned to these two levels are the positive or the negative of the square root of the value given by (23). We have neglected the level zero because it never occurs if $\mathscr{I}$ is even, while, if $\mathscr{I}$ is odd, it occurs only for a very small fraction of the storage elements. In this last case, we can assign one of the two levels, at random, to the zero elements thus causing a negligible increase in the recall variance. After this quantization, if only these two levels are permitted, the memory can only be read. Indeed, no more single memory traces can now be added to the total memory.

The average information stored is the same as in the case considered in Sect. 4.2. When the storage elements are made binary, the upper bound for the storage capacity is then approximately $g^2 0.3$ bit/element. For such a memory system, with one-bit storage elements, $g^2 0.3$ is also the upper bound for the storage efficiency, i.e. this number equals the ratio of the maximum amount of information storable to the capacity theoretically available.

4.4 The Best Strategy to Reduce the Request for Levels

However, if we are interested in drastically reducing the number of bits required by each storage element, but nevertheless preserving the memory system's property of always allowing the storage of new items, then it is necessary to resort to noise keys which are different from those defined in Sect. 2, which have no zero elements. Thus, the use of noise keys which have only a small fraction of their N elements with a non zero value implies that, for each storage process, the storage elements undergo less modification in terms of jumps of levels, hence a smaller number of levels can be sufficient. This case will be discussed in detail in a successive paper.

5 Some Examples

In Table 1 the values of the quantities so far defined to describe the behaviour of the model are reported for a number of cases. For each of three possible choices of the value of $\mathcal{I}/N$, four sub-cases are given. The memory model is studied when the stored memory traces are 10^3 and also 10^6, and, for each of these two cases, when threshold Th is set equal to 0.5 and 0.7. The items are chosen so as to have the same number N of elements as the memory vector, and the recall keys are assumed to be complete, i.e. coinciding with the coding keys used on storage. We then see how the storage capacity of the model changes when the items have less elements than the memory vector or when the keys are incomplete.

The numbers reported in Table 1 confirm the general trend of the model performances. For fixed values of K and Th, an increasing value of $\mathcal{I}/N$ causes H_{TOT} to increase, and the recall efficiency η_R to decrease. The larger the $\mathcal{I}/N$, the larger the amount of information stored per memory element, but the greater the number of errors in the recall. However, for $\mathcal{I}/N$ equal to 0.25, the greatest value considered, the average information stored is very close to the maximum value, so that it will not increase appreciably for greater values of $\mathscr{I}/N$. For fixed $\mathscr{I}/N$ and Th, an increase in the number K of the memory traces causes H_{TOT} to become increasingly close to its upper bound H^{∞}_{TOT}, while η_R decreases. Finally, for our $\mathscr{I}/N$ and K, taking Th = 0.7 improves both H_{TOT} and η_R. This is because, if $\mu<0.5$, i.e. an item has more 0's than 1's, raising the threshold to a suitable value above 0.5, i.e. further from zero, reduces the number of errors in the recall.

The last four columns of Table 1 concern both the case in which the memory traces are singly quantized before storage, as described in Sect. 4.2, and the case in which the total memory is quantized into two levels only after all the items have been stored (see Sect. 4.3). After quantization, the new error probabilities $\hat{\mathscr{P}}_1$ and $\hat{\mathscr{P}}_2$ are greater than the old $\mathscr{P}_1$ and $\mathscr{P}_2$, and, as a consequence, $\hat{H}_{TOT}$ and $\hat{H}^{\infty}_{TOT}$ are smaller than H_{TOT} and H^{∞}_{TOT}, respectively. The current recall efficiency $\hat{\eta}_R$ (not reported here) will accordingly be reduced by the factor $\hat{H}_{TOT}/H_{TOT}$. If the total memory is quantized at the end, then the numbers reported in the last two columns also coincide with the storage efficiency of this read-only memory system, which has binary storage elements.

When the items have a dimension N_0 smaller than dimension N of the memory, while μ, now equal to $(\mathscr{I}/N_0)/K_0$, takes on the same values as above, Table 1 must be read considering the reported data as referred to the case with $K_0 = K/\lambda$ ($\lambda = N_0/N$) traces, instead of K. Of course, the limit values remain the same.

If the memory is read by recall keys which are incomplete according to 1-g, then, for decreasing values of g (<1), an increasingly small amount of information can be stored. For the values of $\mathscr{I}/N$ given in Table 1 and for Th equal to 0.5g or 0.7g instead of 0.5 or 0.7 (so that the recall error probabilities increase), we have, in fact, that the current value for the various H's is of the same order of magnitude as - but always greater than - the corresponding value in Table 1 multiplied by g^2. For example, for Th = 0.5g, H_{TOT} is bounded, as we have seen, by the value given by (18).

6. Concluding Comments

The main properties of the mathematical model of an associative memory based on convolution and correlation, which was proposed by the author (Bottini, 1980), have been quantitatively investigated. In the case studied, the storage of a single item may produce a change in the state of every storage element. The items are binary with digits 0 or 1. The number of levels permitted for each storage element is large enough to make the superimposition of all the memory traces possible without causing the saturation of the storage elements, except as a rare event.

We have found that the storage capacity of the model, i.e. the maximum amount of information that can be stored, on the average, in each storage element, depends on the density μ of the 1's in the items. The dependence of the storage capacity on the degree of completeness of the recall key, mixed - in the missing part - with extraneous noise, has also been determined. In the most favourable case, in which the recall keys are complete, the storage capacity ranges from approximately 0.5 bit/element, when μ is very close to zero, to about one half of this value when $\mu = 0.5$.

The values assumed by the storage elements are as often positive as negative. If, after the items have been stored, we quantize the storage elements into two levels according to their sign, the upper bound for the storage capacity (that is for μ very close to zero) will be approximately 0.3 bit/element, when the recall keys are complete. The amount of information retained in the sign of the values of the storage elements is therefore still large. After quantization we have a read-only memory system whose storage elements are endowed with one bit. In this case, then, at most thirty percent of the theoretical capacity of a storage element can actually be used to store information.

We have pointed out that there is an analogy between measuring the amount of information retrieved in a single recall and measuring the information transmitted over a noisy channel. One could then question whether the Shannon noisy-channel coding theorem might also find an immediate application to our case. This theorem asserts that a message can be suitably coded for transmission with probabilities of error arbitrarily close to zero, despite the disturbing action of the channel, at a rate arbitrarily close to the channel capacity (Shannon, 1948). For items having as many 1's as 0's, i.e. $\mu \doteq 0.5$, the answer is affirmative in the sense that the Shannon theorem allows us to state that the information could be stored, after a suitable pre-coding, in our associative memory, within the limit given by the storage capacity for $\mu =$ 0.5, so as to be retrieved, item by item, with a recall efficiency arbitrarily close to unity. This is because such a pre-coding of the item would produce a vector which still has the same μ. The cases with values of μ less than 0.5 would make further study necessary. However, the above statement has only a theoretical importance. Indeed, as known, the application of the Shannon theorem would require that an ideal condition, namely an infinite length for the items, were met, if an absolute reliability in the recall is desidered.

On the other hand, as is well known, several error-correcting codes have been proposed for practical applications, to improve the reliability of information transmission. If we are interested in investigating the use of such codes in our case, essentially to improve the recall efficiency, we must however realize that the 1's required by any of these coding procedures for the items being stored, while on the one side permitting the correction of a certain number of errors in the recall, on the other, as they contribute in increasing the recall variance, also increase the probabilities of these errors. A favourable balance of these opposite tendencies should therefore be sought.

Finally, it is to be noted that the restriction to the consideration of only binary items can be removed. For items whose elements can take on many values, our procedure for calculating the storage capacity can, in fact, be readily generalized. The necessary number of thresholds to discriminate the values retrieved have to be first introduced, and then, all the ways in which these retrieved values can have been altered by noise have to be taken into account in order to determine the corresponding error probabilities from which the information retrieved can be measured.

Caption

Table 1. Quantitative description of the performance of the model for some sets of values of the ratio $\mathscr{I}/N$, and of Th, and K. In all cases, the density of the 1's in the binary items is small. The columns whose heading includes the symbol "∞", used as apex, concern the case when K approaches infinity. The symbol "$\wedge$" denotes both the case in which the memory traces are singly quantized before storage and that in which the total memory is quantized into two levels, in just one step, at the end.

$\mathcal{I}/N$	Th	$\mathcal{P}_1$	$\mathcal{P}_2$	K	H_{TOT} bit/elem	η_R	H^{∞}_{TOT} bit/elem	$\hat{\mathcal{P}}_1$	$\hat{\mathcal{P}}_2$	$\hat{H}_{TOT}$ bit/elem	$\hat{H}^{\infty}_{TOT}$ bit/elem
0.010	0.5	$2.9\cdot10^{-7}$	$2.9\cdot10^{-7}$	10^3	0.1786	0.990	0.2173	$3.3\cdot10^{-5}$	$3.3\cdot10^{-5}$	0.1468	0.1488
				10^6	0.2171	0.775				0.1488	
0.010	0.7	$1.3\cdot10^{-12}$	$1.3\cdot10^{-3}$	10^3	0.1802	0.998	0.3944	$1.2\cdot10^{-8}$	$8.3\cdot10^{-3}$	0.1783	0.2606
				10^6	0.2797	0.998				0.2558	
0.028	0.5	$1.3\cdot10^{-3}$	$1.3\cdot10^{-3}$	10^3	0.2636	0.573	0.2640	$8.3\cdot10^{-3}$	$8.3\cdot10^{-3}$	0.1882	0.1883
				10^6	0.2640	0.358				0.1883	
0.028	0.7	$1.3\cdot10^{-4}$	$3.6\cdot10^{-2}$	10^3	0.4024	0.873	0.4274	$4.0\cdot10^{-4}$	$7.5\cdot10^{-2}$	0.2778	0.2789
				10^6	0.4274	0.580				0.2789	
0.250	0.5	$1.6\cdot10^{-1}$	$1.6\cdot10^{-1}$	10^3	0.4106	0.120	0.4108	$2.1\cdot10^{-1}$	$2.1\cdot10^{-1}$	0.2716	0.2717
				10^6	0.4108	0.070				0.2717	
0.250	0.7	$8.1\cdot10^{-2}$	$2.7\cdot10^{-1}$	10^3	0.4549	0.136	0.4551	$1.3\cdot10^{-1}$	$3.2\cdot10^{-1}$	0.2905	0.2907
				10^6	0.4551	0.078				0.2907	

TABLE 1

References

Borsellino A, Poggio T (1972) Holographic aspects of temporal memory and optomotor responses. Kybernetik 10: 58-60

Borsellino A, Poggio T (1973) Convolution and correlation algebras. Kybernetik 13: 113-122

Bottini S (1977) Modellistica per interpretare la memoria biologica. Thesis in Physics, Univ. of Padova

Bottini S (1979) Modello matematico di una memoria associativa. Atti del 5° Congresso Nazionale di Cibernetica e Biofisica (Pisa): 335-342

Bottini S (1980) An algebraic model of an associative noise-like coding memory. Biol Cybern 36: 221-228

Bottini S, March R (1985) On a nonlinear associative memory. Internal Report B4-21 IEI CNR (Pisa)

Gabor D (1968a) Holographic model of temporal recall. Nature (London) 217: 548

Gabor D (1968b) Improved holographic model of temporal recall. Nature (London) 217: 1288-1289

Gabor D (1969) Associative holographic memories. IBM J Res Dev 13: 156-159

Kohonen T (1972) Correlation matrix memories. IEEE Trans on Computers C-21: 353-359

Kohonen T (1977) Associative memory. Springer, Berlin Heidelberg New York

Longuet-Higgins HC, Willshaw DJ, Buneman OP (1970) Theories of associative recall. Quarterly Reviews of Biophysics 3: 223-244

Palm G (1980) On associative memory. Biol Cybern 36: 19-31

Shannon C E (1948) A mathematical theory of communication. The Bell System Tech Journal XXVII: 379-423

Steinbuch K (1961) Die lernmatrix. Kybernetik 1: 36-45

Willshaw DJ, Buneman OP, Longuet-Higgins HC (1969) Non-holographic associative memory. Nature (London) 222: 960-962