Chaotic vectors and a proposal for multidimensional associative network

Javed I. Khan and D. Y. Y. Yun

Laboratories of Intelligent and Parallel Systems
Department of Electrical Engineering
492 Holmes Hall, 2540 Dole Street
University of Hawaii at Manoa, HI-96822
javed@wiliki.eng.hawaii.edu

ABSTRACT

The paper presents mathematical and empirical results on the behavior of a new multidimensional neural computing paradigm called **multidimensional holographic associative computing (MHAC)**. MHAC can be potentially used for **high density associative storage** and **retrieval** of image information. Unlike conventional neural computing, each morsel of information in MHAC is represented as a complex vector in a *multidimensional unit spherical space*. Each of the individual phases of the vector enumerates a *value* of the information. The magnitude of the vector represents the associated *confidence* in the information. In contrast, the conventional neural computing operates only on the notion of confidence. The proposed multidimensional generalization demonstrates significant improvement in associative storage capacity without the loss of generalization space. Virtually, unlimited pattern associations can be enfolded over a single holographic memory substrate by higher order encoding. In addition, its well-structured computation, simultaneous muti-channel learning, and single step non-iterative retrieval promise highly scalable parallelism. The paper presents the theory of operation of MHAC that is founded on the generalized holographic principles and multidimensional Hebbian learning. The paper also presents analytical as well as empirical evidence from computer simulation supporting the superior performance of MHAC cells.

1 INTRODUCTION

Associative computing is destined to play a key role in the rapidly evolving field of intelligent image data-base management. Associative computing forms the heart of applications involving content-based retrieval, query by example, indexing by color, texture, shape and real-time database query.

Associative computing has long been thought as an integral part of human cognizance. Optical Holography is one of the first artificial system to demonstrate associative processing capabilities. Since it's advancement, a number of authors[4,5] have drawn the analogy between the brain and distributed memory. In search of simpler representations of optical holographic models of memory,

0-8194-1480-8/94/$6.00

that would be suitable for conventional computer implementation, Gabor[2] noted that systems computing cross-correlation or convolution can mimic Fourier hologram. During the early seventies Willshaw performed comparative studies on various digital associative models such as correlograph and cross bar associative network[6]. Later on more variants such as Hopfield network, Brain State in a Box network, Bi-directional Associative Memory (BAM), etc evolved[1].

Most of these current models of associative memories suffer from two critical setbacks. The most serious of them is their extremely limited storage capacity. The other one is the complication in storing patterns that are very similar to each other, requiring enfoldable patterns to be orthogonal or dissimilar.

In an effort to overcome these setbacks, in this paper we present a new multidimensional associative computing paradigm, which has the potential of challenging both of the aforementioned barriers. The basic computational model is simple and based on pre-existing ideas. In a broad sense, it can be considered as a logical generalization of Hopfield's Network as well as the generalization of optical holographic principles.

In this model, unlike, the conventional artificial neural networks (ANN), each element of information is represented as a *multidimensional complex number*. Conventional ANNs use only scalar representation of information. Previously, Sutherland[3] proposed a 2-D model of network. Interestingly, the original optical holograms also process information in 2-D form as represented through light wave. However, the pioneering efforts simplified the computation mode into scalar. In this research we attempt to investigate the re-institution of the lost dimension. In fact we investigated a more general, multidimensional network with encouraging results. The 2-D representation itself offered higher capacity and retrieval accuracy than the conventional networks. The muti-dimensional generalization seems to even surpass the performance of the later under certain conditions. In addition to higher capacity, the proposed model does not require the enfoldable associations to be orthogonal.

This paper presents mathematical and empirical results on the behavior of this generalized multidimensional holographic network. Section 2 first presents the basic model. Then, section 3 explains the mathematical basis of multi-dimension extension and finally section 4 presents empirical results from extensive computer simulation of this model.

2 THE MODEL

2.1 Representation

A stimulus pattern is represented as a suit of stimulus elements in the form of $S\{s_1, s_2, \ldots . s_n\}$. In conventional ANN, the individual elements or *information* content of a stimulus pattern are expressed and processed as a scalar valued real number extending over a given range. In our

approach, instead, through a suitable transformation, the scalar stimulus pattern elements are mapped onto complex valued numbers with individual dimension elements in the range of $\pi \geq \theta \geq \pi$, such that:

$$s_k \Rightarrow \lambda_k e^{\left(\sum_j^d i_j \theta_j^k\right)}$$

where, $s(\lambda_k, \theta_1^k, \theta_2^k, \ldots \theta_{d-1}^k)$ is a vector which is expanded inside a unit sphere in a d-dimensional spherical space. Each of the θ_j^k is the spherical projection (or phase component) of the vector along the dimension i_j expressing the *content* of information, and λ_k is the magnitude of the vector, expressing the *confidence* on the information inscribed in the phase components.

Sutherland's representation[3] is a direct 2-D special case of this more general representation scheme. The multidimensional mapping of stimulus element from the external scalar field intensities is performed by some non-linear mapping. In many physical cases, the phase components can be directly obtained from the sensors. For, example for color images, the three basic color intensities can be directly translated into three dimensional components resulting in a 4-D representation.

Thus, a stimulus pattern is represented as:

$$[S] = \left[\lambda_1 e^{\left(\sum_j^d i_j \theta_j^1\right)}, \lambda_2 e^{\left(\sum_j^d i_j \theta_j^2\right)}, \ldots . \lambda_n e^{\left(\sum_j^d i_j \theta_j^n\right)} \right]$$

Similarly, the response patterns are also obtained. A similar mapping on the external scalar response field intensities provides the response representation:

$$[R] = \left[\gamma_1 e^{\left(\sum_j^d i_j \phi_j^1\right)}, \gamma_2 e^{\left(\sum_j^d i_j \phi_j^2\right)}, \ldots . \gamma_m e^{\left(\sum_j^d i_j \phi_j^n\right)} \right]$$

2.2 Encoding

In the encoding process, the association between each individual stimulus and its corresponding response are defined in the form of a correlation matrix by the inner product of the conjugate transpose of the stimulus and the response vectors:

$$[X] = [\overline{S}]^T \cdot [R] \qquad \ldots.(1)$$

If the stimulus is a pattern with n elements and the response is a pattern with m elements, then $[X]$ is a $n \times m$ matrix with d-dimensional complex elements.

The strength of this new paradigm is derived from the fact that an enormous number of such associations can be learned and stored on the same $n \times m$ space by superimposing the individual correlation matrices onto the same storage elements.

A suit of associations derived from a set of stimulus and corresponding response is stored in the following correlation matrix X. The resulting memory substrate containing the correlation matrix is referred as Holograph.

$$[X] = \sum_{\mu}^{P} [X^{\mu}] = \sum_{\mu}^{P} [\overline{S}^{\mu}]^{T} [R^{\mu}] \qquad \ldots.(2)$$

2.3 Retrieval

During recall, an excitory stimulus pattern $[S^e]$ is obtained from the query pattern:

$$[S^e] = \left[\lambda_{1,e} e^{\left(\sum_j^d i_j \theta_j^{1,e}\right)}, \lambda_{2,e} e^{\left(\sum_j^d i_j \theta_j^{2,e}\right)}, \ldots . \lambda_{n,e} e^{\left(\sum_j^d i_j \theta_j^{n,e}\right)} \right]$$

The decoding operation is performed by computing the inner product of the excitory stimulus and the correlation matrix X:

$$[R^e] = \frac{1}{N} [S^e] \cdot [X] \qquad \ldots.(3)$$

Where, N is a normalization variable, representative of the number of elements in each stimulus pattern.

If, the excitory stimulus $[S^e]$, bears similarity to any priory encoded stimulus $[S^t]$, then the generated response $[R^e]$ also resembles its corresponding response pattern $[R^t]$. On the other hand, if $[S^e]$ does not correspond to any of the enfolded associations then the elements of $[R^e]$ demonstrates distinctive low magnitude indicating absence of the requested *information* in its enfolded memory.

By combining, the encoding and decoding operations expressed in equation-2 and equation-3, the retrieved association can be decomposed into principal and cross-talk components.

$$[R^e] = \frac{1}{N} \cdot [S^e] [\overline{S}^t]^T [R^t] + \frac{1}{N} \cdot \sum_{\mu \neq t}^{P} [S^e] [\overline{S}^{\mu}]^T [R^{\mu}]$$

$$= [R^e_{principal}] + [R^e_{crosstalk}] \qquad \ldots.(4)$$

Equation-4 provides the following important insight about the characteristics and capability of this new network.

1. If $e \rightarrow t$, then $[R^{e}_{principal}] \rightarrow [R^{t}]$, on the other hand, the cross talk component $[R^{e}_{crosstalk}] \rightarrow 0$, as it appear to be the sum of randomly oriented vectors in d-dimensional space.

2. This new paradigm does not suffer from the linear separability problem like the conventional scalar ANN. The magnitude of the $[R^{e}_{principal}] \rightarrow [R^{t}]$, only when the encoded "information" or the phase components are similar. The relative orientation of linearly non-seperable stimulus pattern vector can not produce spurious minima.

3. The growth of the cross talk component decides the capacity of the network. For, acceptable recall performance, the magnitude of the cross talk must remain well below unity. The magnitude of the cross-talk component indicates the saturation level of the holograph.

3 MATHEMATICAL BASIS FOR MULTIDIMENSIONAL COMPUTING

The saturation or cross-talk component of Holographic substrate as shown by equation-4 can be rewritten in the following form, where A_{μ} represents the resultant product vectors:

$$[R^{e}_{crosstalk}] = \frac{1}{N} \cdot \left\{ \sum_{\mu \neq t}^{P} \vec{A}_{\mu} \right\} \qquad \ldots.(5)$$

For practical purpose, the saturation can be thought as proportional to the sum of a set of randomly oriented vectors. In this section, it will be demonstrated that a set of superimposed unit vectors distributed in m-dimesional space tends to be larger than the superimposition of the same set of vectors performed in an n-dimensional space, when m < n resulting in lower holograph saturation. The tendency increases with the increase in the skew of phase distributions of the vectors.

Equation-5 also shows that the saturation is inversely proportional to N, the length of the stimulus vectors.

3.1 Vectors in Multidimensional Space

Let us consider a set of P vectors with uni-normal projection in n and m dimensional spaces $\vec{A}^{n}_{i}$ and $\vec{A}^{m}_{i}$ respectively, where $1 \leq i \leq P$. In this notation, the number with the vector arrow indicates the dimensionality of the vector. We would like to investigate the following inequality:

$$\left| \sum_{i}^{p} \left[\vec{A}^{m}_{i} \right] \right| \geq \left| \sum_{i}^{p} \left[\vec{A}^{n}_{i} \right] \right| \quad \text{,when } m < n \qquad \ldots.(6)$$

Let the n-dimensional vector (left hand side components of (6)) is represented as:

$$\vec{A}^{n}_{i} \equiv A(\theta^{i}_{1}, \theta^{i}_{2}, \ldots \theta^{i}_{n-1}) \equiv A(x^{i}_{1}, x^{i}_{2}, \ldots x^{i}_{n})$$

Where, the middle representation is the representation of A in n-dimensional spherical co-ordinates and the right representation is its corresponding n-dimensional rectilinear representation. Let, the transformation between the two co-ordinate systems are given by:

$$x_k^i = \prod_{j=1}^{k-2} \cos\theta_j^i \sin\theta_{k-1}^i \qquad \text{when, } k < n$$

$$= \prod_{j=1}^{k-1} \cos\theta_j^i \qquad \text{when, } k = n \qquad \ldots.(7a)$$

Now, let us define a m-dimensional spherical sub-space of the above n-dimensional space using its first m-1 dimensions. Then the uni-normal projection vectors of A in m-dimensional space ($\vec{A}^{m}$), (right hand side components of (6)) become:

$$\vec{A}_i^{m} \equiv A(\theta_1^i, \theta_2^i, \ldots \theta_{m-1}^i)$$

and let its corresponding rectilinear coordinates are:

$$\equiv A(y_1^i, y_2^i, \ldots y_m^i)$$

Where,

$$y_k^i = \prod_{j=1}^{k-2} \cos\theta_j^i \sin\theta_{k-1}^i \qquad \text{when, } k < m$$

$$= \prod_{j=1}^{k-1} \cos\theta_j^i \qquad \text{when, } k = m \qquad \ldots.(7b)$$

3.2 Dimension Dispersion Component

$$\left| \sum_i^P \left[\vec{A}_i^{n} \right] \right|^2 = \sum_k^n \left(\sum_i^P x_k^i \right)^2$$

$$= \sum_k^{n-1} \left(\sum_i^P x_k^i \right)^2 + \left(\sum_i^P x_n^i \right)^2$$

according to equation 7a,

$$= \sum_k^{n-1} \left(\sum_i^P x_k^i \right)^2 + \left(\sum_i^P \prod_k^{n-1} \cos\theta_k^i \right)^2 \qquad (step\,1)$$

$$= \sum_k^{n-2} \left(\sum_i^P x_k^i \right)^2 + \left(\sum_i^P \prod_k^{n-2} \cos\theta_k^i \right)^2 - B_{n-1} \qquad (step\,2)$$

Using (7a) and (7b) and some trigonometric manipulation step 2 has been derived from step 1.

where,

$$B_{n-1} = \sum_{i}^{P} \sum_{j \neq i}^{P} \left\{ [1 - \cos(\theta_{n-1}^{i} - \theta_{n-1}^{j})] \prod_{k}^{n-2} \cos\theta_{k}^{i} \cos\theta_{k}^{j} \right\}$$

In a similar way it can be further derived:

$$| \sum_{i}^{P} \left[\vec{A}_{i}^{n} \right] |^2 = \sum_{k}^{n-3} \left(\sum_{i}^{P} x_{k}^{i} \right)^2 + \left(\sum_{i}^{P} \prod_{k}^{n-3} \cos\theta_{k}^{i} \right)^2 - B_{n-2} - B_{n-1}$$

$$= \sum_{k}^{m-1} \left(\sum_{i}^{P} x_{k}^{i} \right)^2 + \left(\sum_{i}^{P} \prod_{k}^{m-1} \cos\theta_{k}^{i} \right)^2 - B_{m} \ldots\ldots - B_{n-2} - B_{n-1}$$

Using equation 7a,

$$= \sum_{k}^{m-1} \left(\sum_{i}^{P} \prod_{j}^{(k-2)} \cos\theta_{j}^{i} \sin\theta_{(k-1)}^{i} \right)^2 + \left(\sum_{i}^{P} \prod_{k}^{m-1} \cos\theta_{k}^{i} \right)^2 - B_{m} \ldots\ldots - B_{n-2} - B_{n-1}$$

Using equation 7b,

$$= \sum_{k}^{m-1} \left(\sum_{i}^{P} y_{k}^{i} \right)^2 + \left(\sum_{i}^{P} y_{m}^{i} \right)^2 - B_{m} \ldots\ldots - B_{n-2} - B_{n-1}$$

$$= \sum_{k}^{m} \left(\sum_{i}^{P} y_{k}^{i} \right)^2 - B_{m} \ldots\ldots - B_{n-2} - B_{n-1}$$

$$= | \sum_{i}^{P} \left[\vec{A}_{i}^{m} \right] |^2 - \sum_{k=m}^{n-1} B_{k} \qquad \ldots.(8)$$

Each of the B_k term accounts for the change in saturation for adding a dimension and we call it *Dimension Dispersion Component* (DDC) of the vector magnitude.

3.3 Holograph Saturation and Dimension Dispersion Component

DDC plays the key role in the saturation of the holograph. As demonstrated by equation-8, positive DDC asserts the validity of the inequality stated in equation-6. The greater is the value of DDC, the lower will be the saturation in the holographic substrate.

DDC is a function of the distribution of the individual phase components. We would now like to investigate how the distribution of phase may effect the magnitude and in particular the sign of DDC.

Stocastically, the expected value of DDC is:

$$E\{B_n\} = \sum_{i}^{P} \sum_{j \neq i}^{P} E\left\{ [1 - \cos(\theta_n^i - \theta_n^j)] \prod_{k}^{n-1} \cos\theta_k^i \cos\theta_k^j \right\}$$

Assuming statistical independence between the phase components,

$$= \sum_{i}^{P} \sum_{j \neq i}^{P} \left\{ E[1 - \cos(\theta_n^i - \theta_n^j)] \prod_{k}^{n-1} E[\cos\theta_k^i] E[\cos\theta_k^j] \right\}$$

the first term $E[1 - \cos(\theta_n^i - \theta_n^j)] = Q$ is always a positive quantity such that $0 \leq Q \leq 2$,

On the other hand, further assuming uniformity of distribution among the phase components,

$$= {}^{P}c_2 Q \mid E[\cos\theta] \mid^2$$

$$\geq 0$$

From the above analysis we may draw the following important conclusions about the nature of the DDC terms.

(i) Stochastically, the expected value of the DDC terms, in general, will always be positive, irrespective of their distribution.

(ii) Deterministically, if the phase components are distributed in such a way that $-\pi/2 \leq \theta \leq \pi/2$, then also the DDC terms, in general will always be positive.

(iii) The expected value of DDC $E\{B_n\} \rightarrow 0$, as the distribution $f(\theta) \rightarrow C$ implies, $\int_{-\pi}^{+\pi} \cos\theta f(\theta) d\theta \rightarrow 0$. Thus, the lowering of holograph saturation due to dimension dispersion becomes increasingly pronounced as the distribution shifts towards non-uniformity.

(iv) $\cos\theta \leq 1$, hence, the DDC decrease with the increase in dimension. At higher dimension more and more $\cos\theta$ terms contribute to the product. Thus, the improvement due to DDC is relatively more effective at lower range of dimension.

4 EXPERIMENTS

We have performed a set of experiments to empirically measure the improvement due to dimension dispersion. The first two plots, Fig-1(a) and 1(b), show the growth of correlation magnitude (y axis) or the saturation level on the holographic substrate with the increase in the number of stored vectors (x axis). Two figures show the saturation for two different distributions of the stored vectors. To investigate the effect of skew in the stimulus data distribution (expected in natural image data), we have used band limited distributions. The narrower band range represents more skewed stimulus distribution. In Fig-1(a), the phase components are uniformly distributed with in

the range of $+\pi/2 \geq \theta \geq -\pi/2$. On the other hand, in Fig-1(b), these are distributed in the range of $+2\pi/3 \geq \theta \geq -2\pi/3$. For this experiment, the vector elements and their phase components are randomly generated.

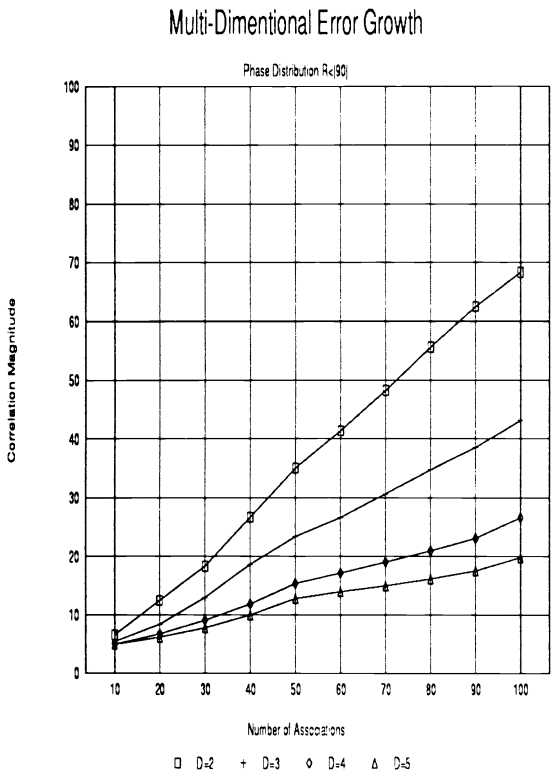

Fig-1(a)

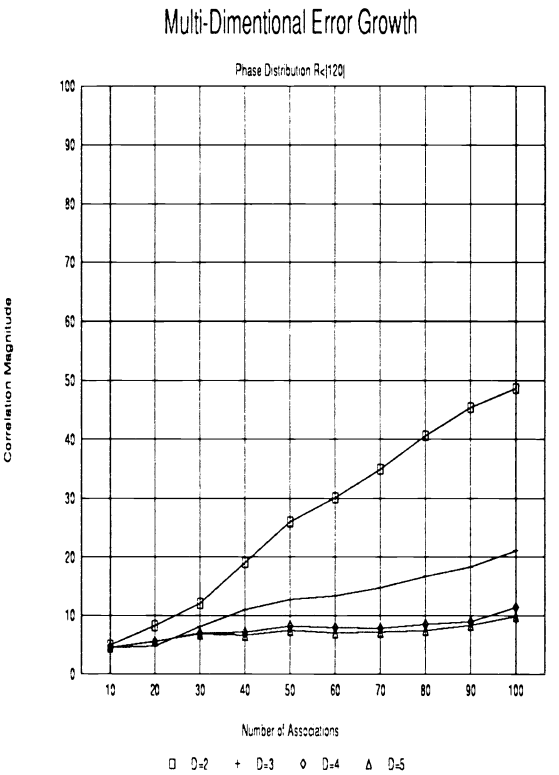

Fig-1(b)

As shown on both of the figures, the saturation increases almost linearly with the increase of number of associations. In general the growth rate is smaller for higher dimensionality. This characteristics is most apparent in Fig-1(a) and less apparent in Fig-1(b). With a limit of operating saturation at 20% (20 in y scale), the number of patterns that can be enfolded with in this limit, increases from 32 for 2-D to 100 for 5-D orientation of the pattern associations in Fig-1(a). With a limit of 10% saturation, the number of enfoldable pattern increase more dramatically, from 25 for 2-D to 100 for 4-D.

Fig-2 plots the saturation with respect to the memory dimension (x axis). It plots the saturation characteristics for several distribution bands of the vector phases. As the dimension is increased, the saturation decreases. The best result is obtained when the phase components are distributed over the full range resulting in very little (~10%) saturation. However, as the band becomes narrower the saturation increases. The difference is more at lower dimension. On the other hand the difference in saturation due to narrow band distribution drops at higher dimension. For example, at 7-D spherical space, the saturation with narrow band distribution, $|R| < 90$ approaches to the ideal distribution with $|R| < 180$.

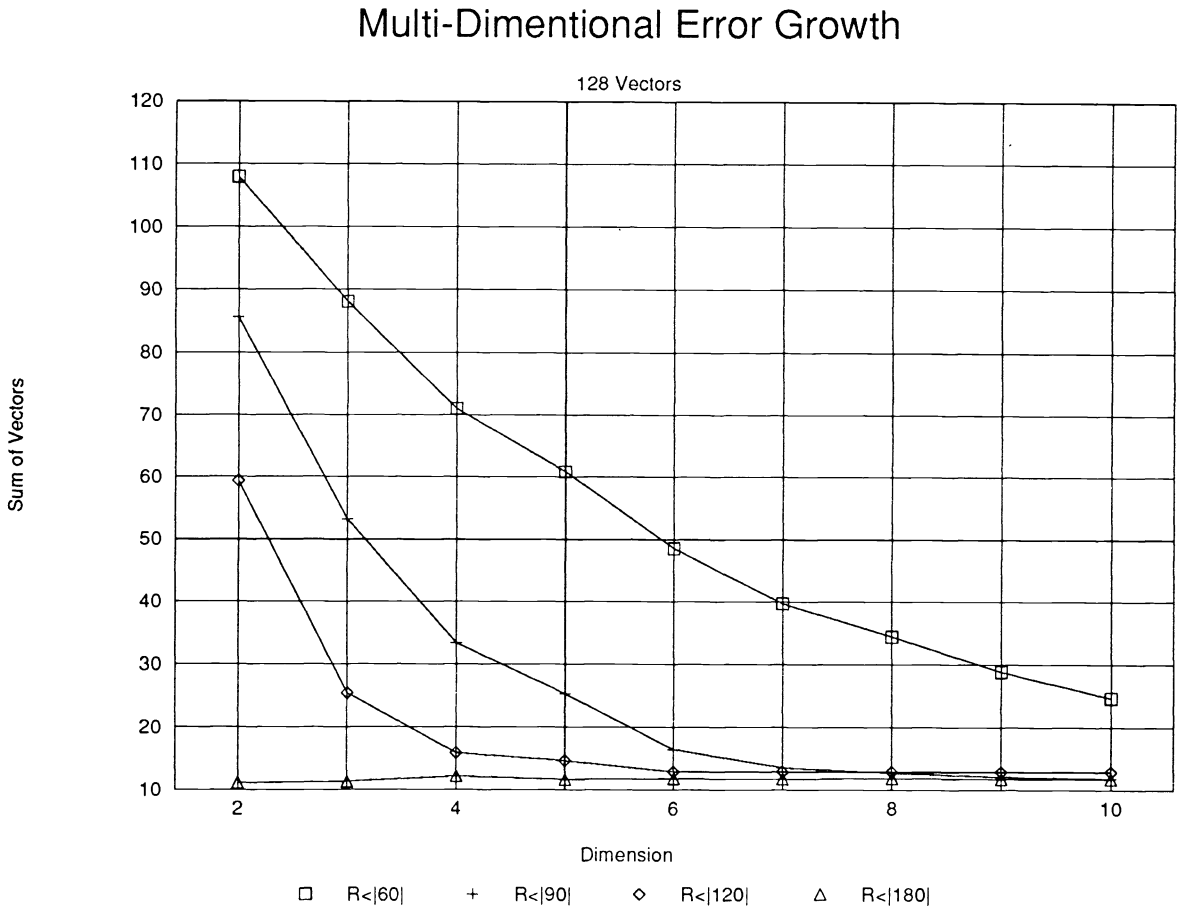

Fig-2

Fig-3(a) and Fig-3(b) finally presents the performance of a simulated single cell multidimensional holographic associative memory. In the test, P (=16) vectors of S (=32) elements, each with D dimensions, are randomly generated. Each of the dimensional phase components are randomly generated with uniform step distribution within the range R.

Each of these stimulus vectors are associated with a matching response vector. The correlation among each of these stimuli and responses are superimposed on a multidimensional holograph. Each of these response vectors, are then recalled using the corresponding stimulus as the query pattern. During the recall process the principal component and the cross talk components of the recalled response have been separately measured. The experiment presents the recall performance in terms of signal-to-noise ratio (SNR) and cross-talk for memories of various dimensions (D=2 to 10).

Fig-3(a) plots the signal to noise ratio (SNR = average signal/ average corss talk) against the dimensionality. Fig-3(b) plots the cross-talk component against dimensionality.

The results indicate SNR improves (Fig-3(a)) and cross-talk (Fig-3(b)) decreases as we use higher dimensional holograph. The improvement is more significant for the narrower band stimulus. For |R| <90, the cross talk component reduces from 13% to 2.5% as the dimensionality is improved from 2-D to 4-D. Thus the improvement due to dimension dispersion is more prominent when the skew in input vector distribution is higher.

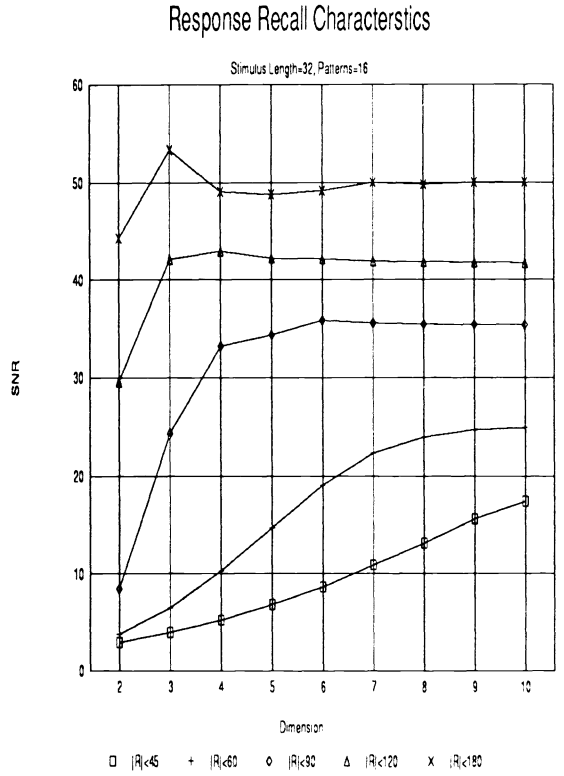

Fig-3(a)

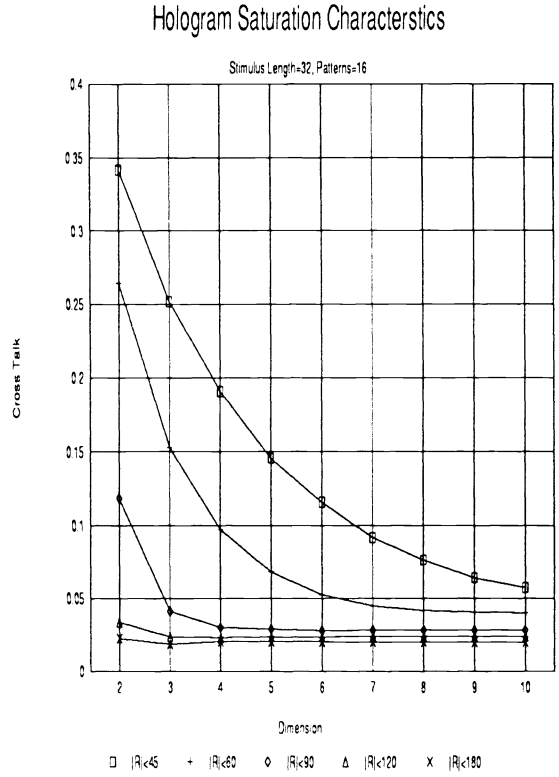

Fig-3(b)

5 CONCLUSION

As shown in equation-5, the capacity of the holograph is also linearly proportional to the length of the stimulus pattern. Experiments have shown that the number of images that can be superimposed on a single holograph and retrieved with less than 5% recovery error is almost .5 times the number of pixels in each image. This capacity is linearly expandable with stimulus length. In fact, higher order encoding can make the capacity virtually unlimited[3].

The results presented in this paper implies that by increasing the memory dimension potentially larger number of associations can be stored in MHAC than conventional associative memories. However, the digital reconstruction of the holograph on conventional digital memory may not directly correspond to the performance gain indicated in our experiment. The actual digital memory requirement also involves the bit allotment to represent individual dimensions of the holograph elements, which in terms is dependant of the distribution and nature of the data.

The encoding process of MHAC (equation-2) shows that all the pattern associations can be stored simultaneously. On the other hand, the decoding process (equation-3) is also non-iterative single step computation. The computations involved in both encoding and decoding are straight forward multidimensional matrix product operations. Thus, the entire operation of MHAC offers high degree of parallelism with a potential of real time application. Interestingly, depending on the application constraints a trade-off can be performed at the encoding stage. Our experiments has shown that a Differential Hebbian learning (DHL), instead of Hebbian learning, can further increase the capacity of MHAC. However, DHL is intrinsically sequential.

6 REFERENCES

1. Caudill, M. & C. Butler, Naturally Intelligent Systems, MIT Press, 1990.

2. Gabor, D., Associative Holographic Memories, IBM Journal of Research and Development, 1969, I3, p156-159.

3. Sutherland, J., Holographic Models of Memory, learning and Expression, International J. of Neural Systems, 1(3), pp356-267, 1990.

4. van Heerden, P. J., A New Optical Method for Storing and retrieving Information, Applied Optics, 1963, 2, pp387-392.

5. Westlake, P. R., Towards a Theory of Brain Functioning: A Detailed Investigation of the Possibilities of Neural Holographic Processes, Doctoral dissertation, University of California, Los Angeles, 1968.

6. Willshaw, D. J., Models of Distributed Associative Memory, Edinburgh University, 1971.