(2)

FINAL REPORT

OPTICAL IMAGE RECOGNITION USING ACOUSTOOPTIC AND CHARGE COUPLED DEVICES

Submitted to:

U.S. ARMY RESEARCH OFFICE
P.O. Box 12211
Research Triangle Park, NC 27709

OPTICAL IMAGE RECOGNITION USING ACOUSTO-OPTIC AND CHARGE COUPLED DEVICES

FINAL REPORT

PROFESSOR DEMETRI PSALTIS

5 MARCH 1989

U.S. ARMY RESEARCH OFFICE

CONTRACT/GRANT NUMBER DAAG29-85-K-0257

California Institute of Technology
Pasadena, California, 91125

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS: N/A
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARO 22985.18-PH-SDI
6a. NAME OF PERFORMING ORGANIZATION: California Institute of Technology; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code): Pasadena, CA 91125	7b. ADDRESS (City, State, and ZIP Code): P. O. Box 12211, Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U. S. Army Research Office; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAAG29-85-K-0257
8c. ADDRESS (City, State, and ZIP Code): P. O. Box 12211, Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Optical Image Recognition using Acoustooptic and Charge Coupled Devices

12. PERSONAL AUTHOR(S)
Demetri Psaltis

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
final	FROM 10/15/85 TO 10/15/88	1989 March 5	

16. SUPPLEMENTARY NOTATION
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Optical Information Processing; Neural Networks
Pattern Recognition; Volume Holography
Image Processing; Photorefractive holograp

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Optical image recognition systems based on optical correlators and distributed holographic interconnections are described and experimenatally demostrated. Correlators which rely on acoustooptic devices, optical disks, photorefractive crystals, and charge coupled devices are described. The capacity for information storage of such correlators is derived. The capacity of volume holograms for high spatial density information storage is derived. The use of photorefractive volume holograms in artificial neural networks is described and demonstrated. Finally, an analysis of the use of binary filters in pattern recognition systems is presented and optical neural networks which rely on higher order correlations are described.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

FOREWORD

This report is composed of reprints of papers published over the course of this program. Each paper deals with some aspect of the design and capacity of optical pattern recognition systems. Paper#1 describes architectures and implementations for optical correlators based on acoustooptic devices. Paper #2 analyses the capacity of optical correlators for image recognition. The third paper describes optical correlators which use optical disks, rather than AO cells, as spatial light modulators. Papers 4 and 5 are concerned with optical image recognition systems based on neural network models. The optical associative memory described in these papers uses a liquid crystal spatial light modulator, thin holographic plates and a CCD to process two dimensional images. Papers 6-9 deal with the use of photorefractive crystals to increase the adaptability and capacity of optical image processing systems. Paper 6 describes a system which uses AO cells and a bismuth silicon oxide crystal to implement a time integrating correlator. Paper 7 explores the use of volume holographic materials, such as photorefractive crystals, as media for storing spatial information at high densities. Various constraints are derived for the storage and reconstruction of holographic information. Papers 8 and 9 describe methods for using the dynamic nature and high storage capacity of photorefractive crystals to construct artificial neural networks. Paper #10 presents an analysis of optical image recognition systems based on binary filters. Paper #11 describes an extension of correlation optical associative memories to higher orders which results in higher storage capacity.

ACOUSTOOPTIC IMAGE CORRELATORS

Sung Gi Paek, Cheol Hoon Park, Pai Mok, and Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California, 91125

Introduction

Two dimensional correlation for performing image recognition is one of the earliest examples of optical information processing suggested by Vander Lugt [1] and it still remains one of the most promising application areas for this technology. The reasons optical image correlators have not found to this day widespread usage in pattern recognition systems fall into two categories: algorithmic limitations and lack of devices. Correlation-based pattern recognition algorithms have well known limitations such as scale and rotational sensitivity. Over the years algorithms based on optical correlation have been developed [2-5], which show promise for overcoming some of the algorithmic limitations, however they have not been put to a real practical test because of the lack of two dimensional spatial light modulators that are necessary for the implementation of the classic Vander Lugt correlator. Specifically, two separate two dimensional SLMs are needed, one for recording the input image and another for the reference.

Recently, there has been considerable progress in the development of spatial light modulators which has led to several "real time" implementations of optical correlators using two dimensional devices [6-8]. In this paper we report on the use of acoustooptic devices(AODs) which are by far the most highly developed SLMs available. AODs however are one dimensional SLMs and in order to process two dimensional signals in an acoustooptic image corelator, we use a combination of temporal and spatial integrations [9-11]. In this paper we briefly review the principles of operation of acoustooptic image correlators and then present the results of several experiments that were performed with real time systems that we have assembled.

In the architectures we will discuss, an AOD is used as the transducer for the input image that is to be recognized. Another aspect we will explore is ways of making the reference image programmable rather than being fixed on a hologram recorded on a photographic film. We will see that time domain processing, an inherent part of the acoustooptic image correlators, offers a very convenient way of introducing a programmable reference and creates a set of promising algorithmic possibilities that are not realizable with a fixed reference.

The general method

Single transducer acoustooptic devices are one dimensional spatial light modulators and hence they can not be used to represent an input image in its entirety at one time. Typically the space bandwidth product of the AOD is about equal to the number of resolvable spots along only one line of the image we want to process. As a result, the strategy for building an acoustooptic image correlator is process the image one line at a time using a space integrating optical processor and accumulate by temporal integration on a two dimensional detector array the results of the space integrating part of the system to form the full two dimensional correlation. The basic idea will be explained with the aid of Figure 1. The operation we wish to perform is a two dimensional correlation:

$$g(x',y') = \iint f(x,y)h^*(x+x',y+y')\,dxdy. \qquad (1)$$

The input image $f(x,y)$ is imaged onto a TV camera that electronically scans the image and produces a video signal. The video signal is heterodyned to the appropriate center frequency and then applied to the piezoelectric transducer of the AOD. Each horizontal video line propagates separately in the AOD and modulates the incident light. Since in our processors the images are processed as individual lines, from now on we will denote the input image as $f(x,nd)$, where n is an integer and d is the line spacing in the vertical direction. The optical system is a multichannel one dimensional correlator that produces the correlation between each input video line in the AOD and all the lines of the reference image which is stored in the optical system. The light incident on the two dimensional CCD detector at the output is modulated by:

$$g_n(x',y') = \int f(x,nd)h^*(x+x',y')\,dx \qquad (2)$$

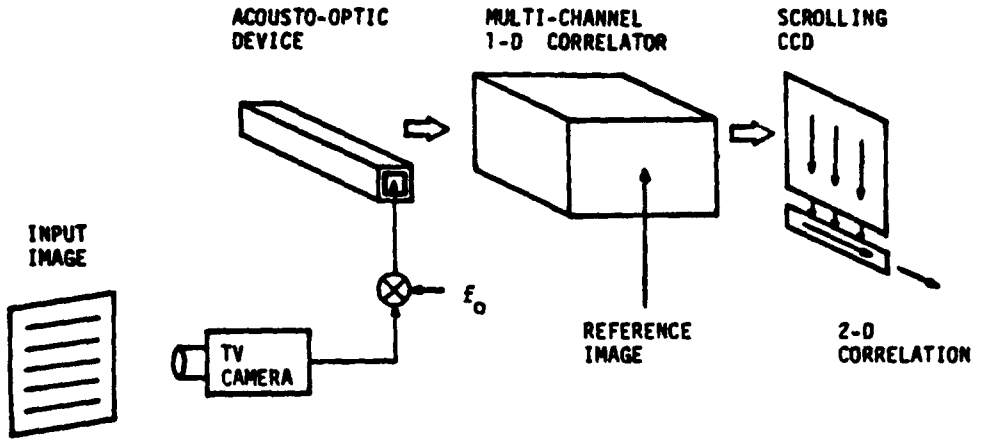


Figure 1. Two dimensional accoustooptic image correlator.

where x' and y' are the spatial coordinates at the output plane. In order to complete the two dimensional operation, we need to shift the two dimensional pattern $g_n(x',y')$ by a distance nd in the y' direction and then sum over n (i.e. accumulate the signal from all the different input lines). The required shift and add operation can be accomplished very conveniently by continually scrolling the photogenerated charge on the CCD during each video frame. After the nth input horiontal line is scanned by the TV camera and the correlation between this nth input line and the reference is added onto the previous contents of the CCD, the CCD is then triggered to electronically transfer the entire charge pattern vertically by one pixel. This procedure results in the formation of a charge pattern on the CCD that can be expressed as follows:

$$g(x',y') = \sum_n g_n(x',y'+nd) = \sum_n \int f(x,nd)h^*(x+x',nd+y')\,dx \qquad (3)$$

The above is recognized as a sampled form of the two dimensional correlation in Eq. (1). The two dimensional correlation is produced continuously at the frame rate of the input TV camera and it appears at the CCD output in the form of a standard video signal that can be directly displayed.

Within this basic framework, there are several possible architectural variations principally through choosing different implementations for the multichannel one dimensional correlator. In references 13 and 14 we have described in detail two specific architectures. In what follows we present results from the experimental demonstration of these systems.

Holographic accoustooptic image correlator

The holographic image correlator is shown in Fig.2. The details of the operating principles have been reported in reference [10]. Here we describe the experimental apparatus. The input image was sensed with a high resolution TV camera. The video signal from the TV camera was heterodyned to the center frequency of the AOD (50 MHz), amplified and applied to the AOD. The accoustooptic device in this experiment was a TeO_2 (Crystal Technology # 4050S) device with 35 MHz bandwidth and 70 microsecond delay. This was more than adequate to accomodate one standard video line (63 microseconds and 5 MHz). After approximately 52.7 microseconds from the start of the horizontal clock the signal in the AOD is an acoustic replica of the video line from the input image. At that instant the laser diode is triggered to produce a short pulse to freeze and read-out the signal in the AOD. The laser diode used in the experiments was RCA C86030E with peak power equal to 40 milliwatts and pulsewidth equal to 50 nanoseconds. The pulsewidth must be chosen equal to or shorter than the inverse of the bandwidth of the video signal so that the motion of the signal in the light diffracted by the AOD can be neglected. The video bandwidth in these experiments was 5MHz or less. The light diffracted by the AOD was Fourier transformed in the horizontal direction and expanded vertically to illuminate a one dimensional Fourier transform hologram of the reference image. The light diffracted by the hologram was imaged vertically and transformed horizontally to produce the multiple one dimensional correlations at the CCD plane.

The hologram was fabricated in dichromated gelatin yielding efficiency 35% at the 820nm wavelength. The intensity of the light was detected at the CCD plane. The horizontal clock

from the CCD is used to trigger the driving electronics of the input TV camera and to transfer the charge in the array downwards by one pixel. The CCD device used in the experiments was a SONY XC-37 whose driving electronics were modified to allow us to scroll the charge on the CCD continuously during each frame. This CCD camera has 384 pixels in the horizontal direction and 491 pixels in the scrolling direction. The scrolling action of the CCD completes the 2-D correlation as described earlier and the full 2-D correlations between the image to which the TV camera is pointed and the reference image stored on the Fourier transform hologram is produced at 30 frames per second and displayed on a monitor.

A photograph of the experimental apparatus is shown in Figure 3 showing that the system is relatively simple and compact. The laser diode is on the right and the CCD on the left side of the photograph in Figure 3.

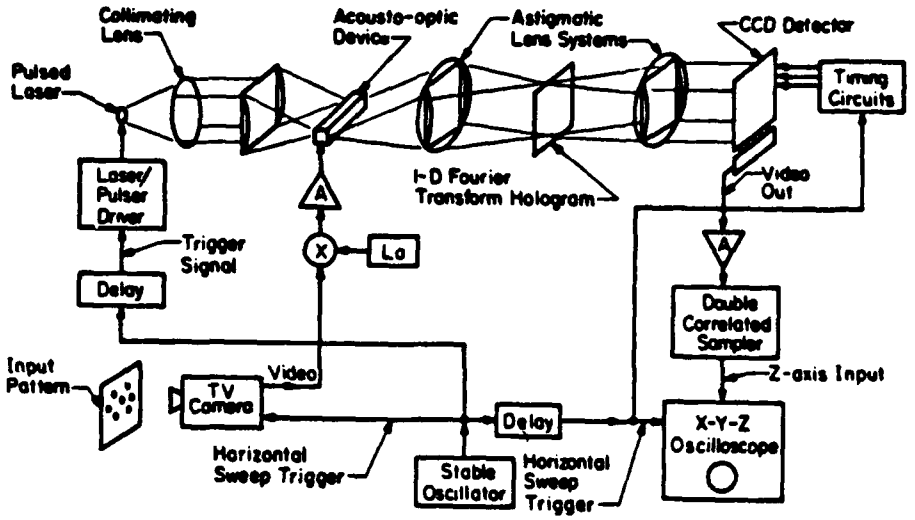


Figure 2. Holographic acoustooptic image correlator.

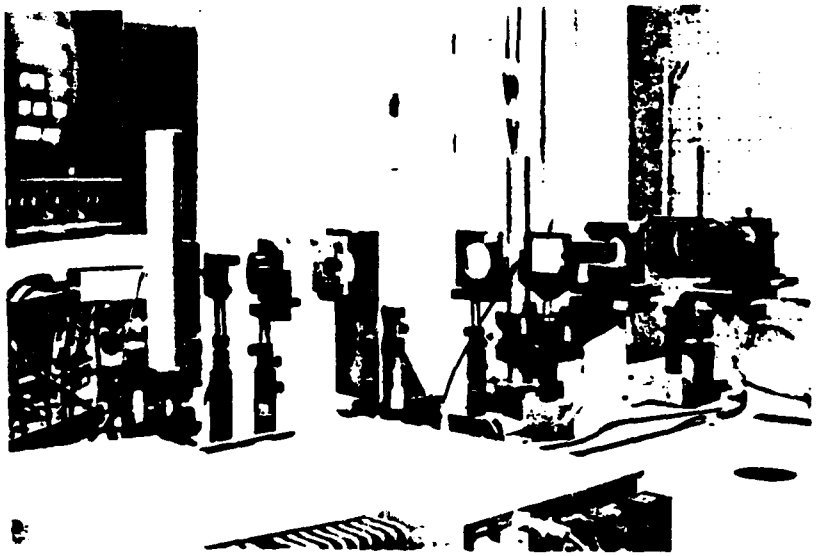


Figure 3. Experimental setup of the system in Figure 2.

An example of the experiments that were done with this system is shown in Figure 4. The two faces shown at the top of Figure 4 were used as inputs. On the left hand side of the photograph the reconstructions of the two faces are displayed. These reconstructions were

obtained by pointing the TV camera to an impulse-like function and thus producing at the output the impulse response of the system. It is evident that the impulse response of the system is an edge enhanced version of the original in order to suppress the crosscorrelations. This was accomplished by fabricating the hologram so that higher frequency components were enhanced. All four possible correlations and their cross sections are shown in Figure 4 in two dimensional display. The results show excellent discrimination between the two faces.

The photograph shown in Figure 5 is an isometric view of the autocorrelation of one of the patterns in Figure 4. A very sharp correlation peak is obtained which is significant because the correlation produced with this system is partially incoherent. In a fully incoherent correlator that operates on light intensity [12], the correlation peak is typically broad and rests on top of a bias plateau. In the system we are describing in this section, the correlation in the horizontal direction is performed by coherent amplitude integration, whereas in the vertical direction it is done by integrating intensity. This incoherent integration averages out the coherent noise effects (i.e. speckle and related phenomena) and yet, the correlation pattern that is produced is bias free and its sharpness in both dimensions is characteristic of coherent correlators.

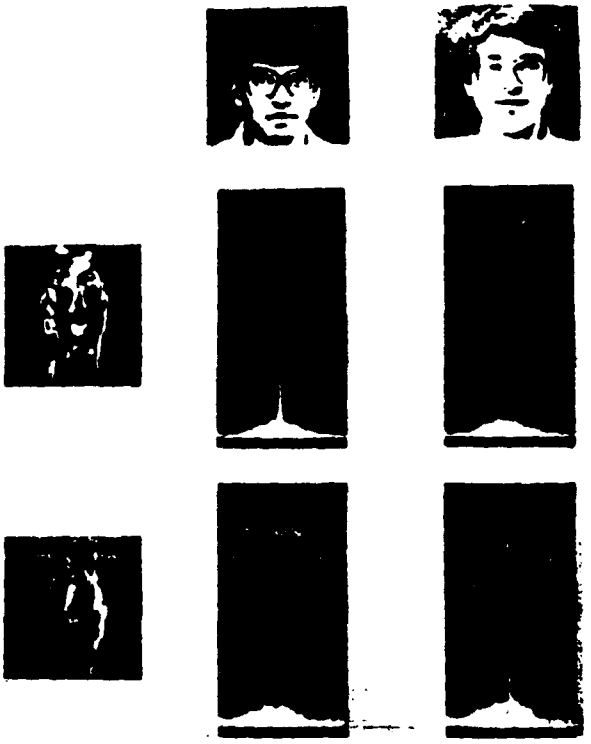


Figure 4. Experimental demonstration of the holographic acoustooptic correlator.

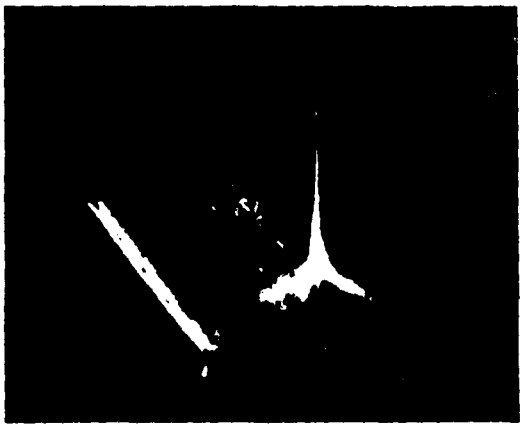


Figure 5. Isometric view of the autocorrelation of one of the patterns in Figure 4.

The results that we obtained from the system described above have convinced us that acoustooptics can essentially solve the device limitations of optical image correlators, at least as far as the real time input stage is concerned. If we ask what are the limitations of this system in terms of performing adequately in a pattern recognition application, the answer is clearly not that the optical system does not perform correlations well enough. The real issue now is how to use an optical correlator to recognize images. We have come to the conclusion that a digitally programmable reference is a key feature that needs to be incorporated in optical correlators in order to make their applicability to pattern recognition practical.

The most obvious way to introduce a programmable reference to the system in Figure 2 is by recording the Fourier transform hologram on a real time SLM rather than on photographic film. We have investigated the use of the Litton magnetooptic spatial light modulator [13]

A photograph of the experimental apparatus is shown in Figure 8. The LED used in the experiment is fabricated by Honeywell and it consists of 180 elements, each separated by 100 microns. Each element transmits 5 microwatts of optical power. In the experiment we will describe, only 32 elements of the array were used. The electronic memory that is needed for storing the reference image was built with standard RAM chips and it was designed to store binary images consisting of 32X100 pixels. The memory was interfaced to the optical systems through an array of 32 LED drivers and to also to an IBM personal computer which was used to generate the reference patterns. The AOD and the CCD were the same devices that were used in the experiments described earlier.

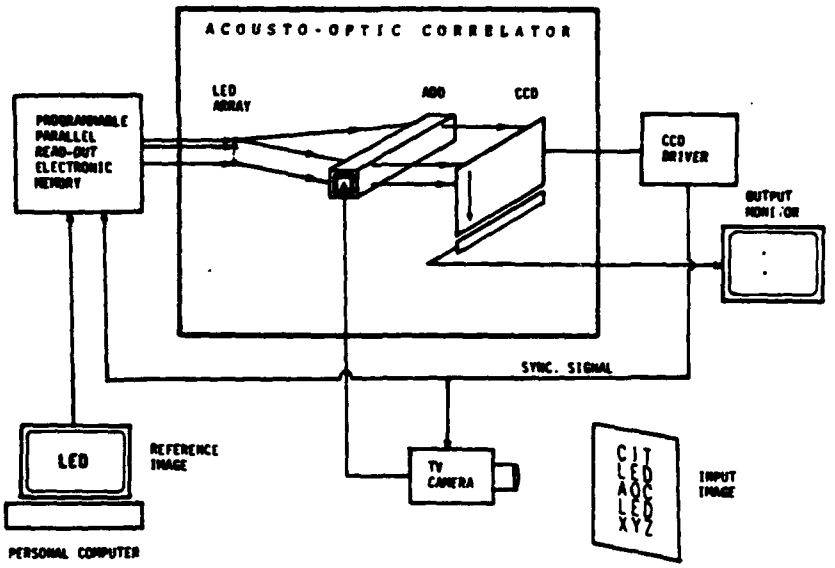


Figure 7. Incoherent LED acoustooptic correlator.

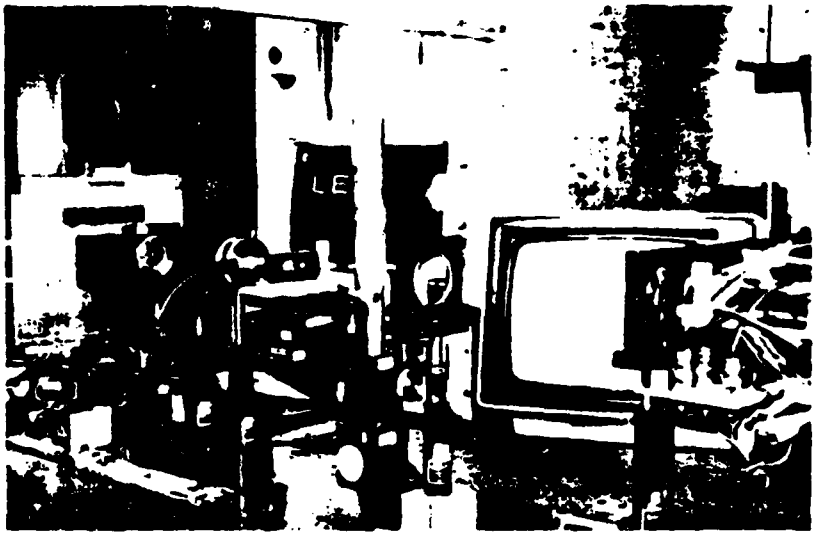
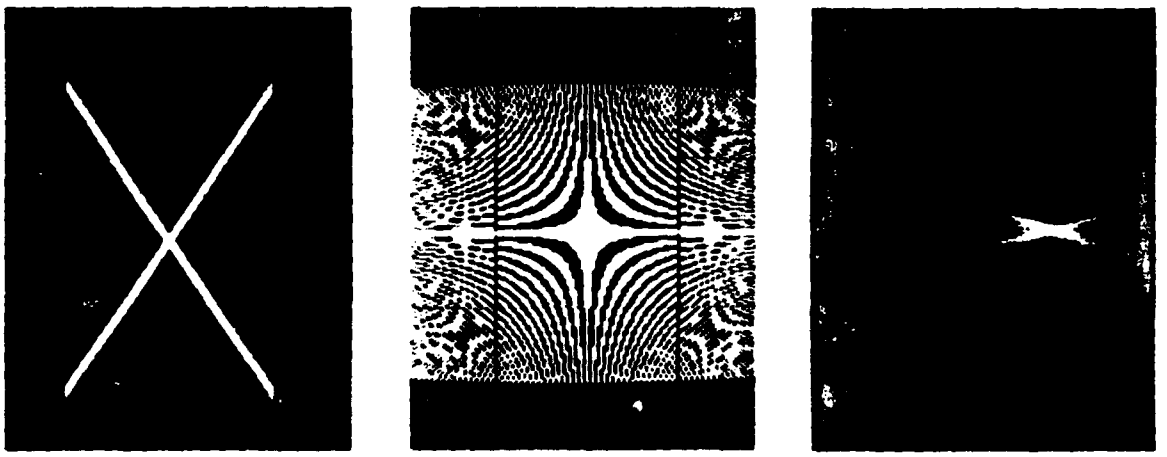


Figure 8. Experimental setup of the system in Figure 7.

A sample of the experimental results obtained with this system are shown in Figure 9. The input pattern (Fig. 9a) contains the word "LED" in two places. The reference image chosen in this case to be the word LED and shown in Figure 9b is the reference image as displayed on the screen of the computer. The output of the optical correlator is shown in Figures 9c and

as a spatial filter in optical correlators [14]. The magnetooptic device (MOD) we are using is structured as a two dimensional array of 128X128 pixels. Each pixel can be electrically set to one of two possible magnetization states. Hence this is a binary spatial light modulator. A feature of the device that is most interesting in terms of recording holograms is the fact that it is a bipolar light modulator [14]. This implies that the average amplitude transmittance of the hologram can be made zero which allows an on-axis Fourier transform hologram to be recorded and read-out. A very simple algorithm for recording computer generated Fourier transform holograms on this device was described and demonstrated in reference [14]. The real part of the Fourier transform of the reference image is examined for each pixel. If it is positive, the corresponding pixel of the MOD is set to one of its two states, and if it is negative, it is set in the opposite state. This algorithm has yielded excellent results in a conventional two dimensional optical correlator. It can be modified for the acoustooptic correlator in Figure 2 by simply calculating the one dimensional Fourier tranform of the reference image with a digital computer, then recording on the MOD the sign of the real part of this transformed image. A preliminary experimental result with a system similar to the one in Figure 2, with an MOD replacing the hologram, was obtained. The results are shown in Figure 6. The input pattern that was imaged onto the TV camera was the letter X shown in Figure 6a. The one dimensional Fourier transform hologram of the letter X, computed and then recorded on the MOD, is shown in Figure 6b. The autocorrelation of the letter X that was produced in real time as a video signal by the CCD and displayed on a monitor, is shown in Figure 6c.



a b c

Figure 6. Experimental demonstration of the holographic acoustooptic correlator with a programmable magnetooptic device. (a) Input. (b) Hologram. (c) Correlation output.

Incoherent LED acoustooptic correlator

In this section we present results from an experimental demonstration of the incoherent correlator that is described in detail in reference [11]. A schematic diagram of this processor is shown in Figure 7. The basic architecture follows from Figure 1. The difference between this and the system discussed in the previous section is the choice of the multichannel one dimensional correlator. The system in Figure 7 utilizes an array of incoherent time integrating correlators rather than coherent space integrating correlators. The temporal signal modulating each of the LEDs in Figure 7 is correlated against the signal that is launched into the acoustooptic device. These correlations are formed on separate lines of the CCD at the output of the system. The reference image is stored in electronic memory which can be read-out in parallel such that each line of the reference image can temporally modulate a separate LED. The electronic memory is triggered to read-out its contents in synchronism with the horizontal clock from the CCD. As each new input video line is entered in the AOD, it is correlated against all the lines of the reference image. As in the previous architecture, the CCD is triggered to scroll in synchronism with the horizontal sync of the input TV camera, which at the end of each video frame, results in the formation of the full two dimensional correlation between the input image and the reference image stored in the electronic memory.

9d. These are pictures of the same thing with the contrast of the display monitor adjusted to display all the cross correlations in figure 9c, whereas in Figure 9d only the autocorrelation peaks are visible. The crosscorrelations in this case are higher, when compared with the holographic correlator. This is because in an incoherent correlator, only positive values can be directly represented. We are currently working on incorporating in this system the capability to represent bipolar images. This permits much greater flexilility in choosing the reference image which in turn permits the selection of the reference image to minimize the crosscorrelations and obtain performance comparable with the coherent correlator.

a b c d

Figure 9. Experimental demonstration of incoherent LED correlator. (a) Input. (b) Reference image. (c),(d) Correlation output.

Acknowledgement

We acknowledge Litton Data Systems for their support and the loan of the magnetooptic device that was used in the experiments. We also acknowledge Mr. Gotoh Goich at the Sony Corporation for his help with the CCD used in the experiments.

The work reported in this paper is funded by grants from the Army Research Office, the Air Office of Scientific Research and General Dynamics.

References

1. A. Vander Lugt, "Signal detection by complex spatial filtering," IT-10, pp.139-145, 1964.
2. H.J.Caulfield, R.Haimes, and D.Casasent, Opt. Eng., 19(2), 152 , 1980.
3. D.Casasent and D.Psaltis, "New optical transforms for pattern recognition," Proc. IEEE, vol.65, pp.77-84, 1977.
4. E.G.Paek and S.S.Lee, "Discrimination enhancement in optical pattern recognition by using a modified matched filter," Can. J. Phys., vol.57, pp.1335-1339, 1979.
5. H.H.Arsenault and Y.N.Hsu, "Rotation-invariant discrimination between almost similar objects," Appl. Opt., vol.22, pp.130-132, 1983.
6. J.G.Duthie and J. Upatnieks, "Compact real-time coherent optical correlators," Opt. Eng., vol.23, pp.7-11, 1984.
7. B.D. Guenther, C.R. Christensen, and J. Upatnieks, "Coherent optical processing : Another approach," IEEE J. of Q.E., vol.QE-15, pp.1348-1362, 1979.
8. A.D.Gara, "Real-time optical correlation of 3-D scenes," Appl. Opt., vol.16, pp.149-153, 1977.
9. D.Psaltis, "Two-dimensional optical processing using one-dimensional input devices," Proc. IEEE, vol.72, pp.962-974, 1984.
10. D.Psaltis, "Optical image correlation using acoustooptic and charge-coupled devices," Appl. Opt., vol.21, pp.491-495, 1982.
11. D.Psaltis, "Incoherent acoustooptic image correlator," Opt. Eng., vol. 23, pp.12-15, 1984.
12. A.W.Lohmann and H.W.Werlich ,"Incoherent matched filtering with Fourier holograms," Appl. Opt., vol. 10, pp. 670-672, 1971.
13. W.E.Ross, D.Psaltis, and R.H.Anderson, Opt. Eng., vol. 22(4), 485, 1983.
14. D.Psaltis, E.G.Paek, and S.S.Venkatesh, "Optical image correlation with a binary spatial light modulator," Opt. Eng., vol.23, pp.698-704, 1984.

Published in "SPIE, Vol.825-22, August, 1987, San Diego, CA"

CAPACITY OF OPTICAL CORRELATORS

Jeff Yu, Fai Mok, and Demetri Psaltis

Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125

ABSTRACT

The capacity of the Vander Lugt correlator, defined as the maximum number of separate images that can be recognised, is estimated. The increase in capacity that results from the use of a volume hologram in place of the commonly used planar hologram is derived. The effects of binarising the reference filter and the shift invariant properties of the two classifying systems are also analysed.

I. Introduction

Vander Lugt correlators have been used for a long time in optical pattern recognition[1]. In the typical implementation, shown in figure 1, the Fourier trasform of an input image is used to read out a hologram containing the Fourier transform of a reference image. This diffracted beam is then inverse Fourier transformed to produce the correlation between the input and reference images on the output plane. Vander Lugt correlators are typically used as pattern recognition systems. Whether or not a peak is present at the output of the correlator determines whether or not the input image is sufficiently close to the stored reference . Recently much work has been done on the use of a Vander Lugt correlator to pattern classifiation [2][3]. In this case, the correlator distinguishes whether the input is a member of one of two classes with each class being composed of many images. Typically, a reference filter is formed as a linear combination of the images in both classes and the presence or absence of a peak at the correlation plane determines which class the input belongs to.

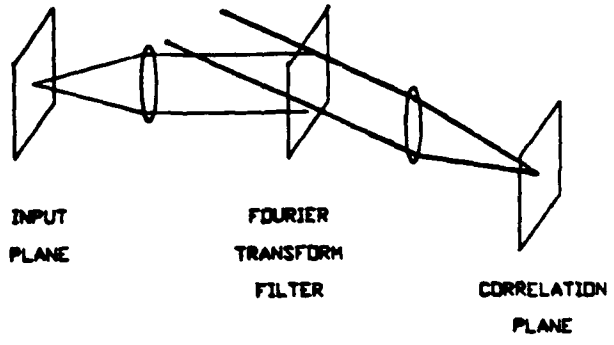


Fig. 1. Vander Lugt correlator.

In this paper, we discuss the capacity of the Vander Lugt correlator. This is to say we estimate the maximum number of images that can be stored in the reference filter before the system begins to misclassify images. This capacity has been studied in great detail for systems without shift invariance (e.g.perceptrons). The classic results from pattern recognition about the capacity of a linear discriminant function do not directly apply in this case because the VanderLugt correlator is shift invariant. In this paper, we will discuss the capacity of the system incorporating the shift invariance of the Vander Lugt correlator. We will also discuss the effect on the capacity of binarizing the reference filter and lastly we will demonstrate that by using a volume holgram to record the filter, the capacity of the system is greatly increased, as well as be becoming capable of multi-class classification.

II.Capacity of Linear Filters

In the most common pattern classification scheme, the inner product is performed between the input image

$\phi(x,y)$ composed of N pixels and a reference filter $h(x,y)$.

$$O = \sum_{y'=1}^{\sqrt{N}} \sum_{x'=1}^{\sqrt{N}} h(x',y')\phi_i(x',y') \quad [1]$$

Comparing the output O with a pre-set threshold, determines which of the two classes the input belonged to. A standard method of forming the reference filter is as a linear combination of the images in both classes:

$$h(x,y) = \sum_{i=1}^{M} w_i \phi_i(x,y) \quad [2]$$

The weights w_i can be chosen through a variety of training algorithms such as the perceptron learning algorithm. It is a well known result that the capacity of such a system is [4]

$$M = 2N \quad [3]$$

where N is the number of pixels in each image. In this paper, we will consider the construction of a simpler filter in which the weights are binary.

$$w_i = \begin{cases} 1 & \text{if } \phi_i \in \text{Class I} \\ 0 & \text{if } \phi_i \in \text{Class II} \end{cases} \quad [4]$$

In other words, the filter is formed by simply summing the images belonging to class 1, while ignoring those in class 2. This is implemented in a Vander Lugt correlator, by multiply exposing the holgram to the images in class 1 while doing nothing for the images in class 2. Classification can then be performed by detecting and thresholding the output at the center of the correlation plane. For the remainder of the paper, we will assume that the images $\phi(x,y)$ consist of binary N pixels, each pixel being a bipolar (ie 1 or -1), independent random variable. Under these assumptions, the capacity of the VanderLugt correlator using the peak-only detection scheme can by found be solving the following transcendental equation[5]:

$$M = \frac{N}{4log(M^3/N)} \quad [5]$$

As $N \rightarrow \infty$, the above expression asymptotically approaches

$$M = \frac{N}{8logN} \quad [6]$$

Thus the use of the simpler method for constructing the reference filter, results in a relatively modest loss in capacity by a factor of 16logN

III. Capacity of Shift Invariant Filters

Because Vander Lugt correlators are inherently shift invariant it is possible to classify prescribed images and their shifted versions as well. In order to implement a shift invariant classification scheme, detection at the output is done over the entire correlation plane. As a result the detection of a peak anywhere in the output plane determines whether the input is a member of class 1 or a shifted form of a member in class 1. Figure 2a shows a cross section through the origin of the digital correlation of an input image with a filter containing only one image. The resulting output shows a single correlation peak and relatively small sidelobes. When the reference is constructed by adding 3 images (figure 2b, the sidelobe structures shows a significant rise in amplitude. However, since only the single correlation peak lies above the threshold, classification of the input image is still performed correctly. However, when the number of reference images is increase to 6 (Fig.2c), there are now two peaks which lie above the threshold level. As a result, the system can no longer decide whether the input image is a member of class 1 or a shifted version of a member of class 1. Therefore, we expect that the capacity of the shift invariant system is smaller. For the relatively simple method of filter construction , we can readily derive an analytic capacity for the shift invariant correlator.

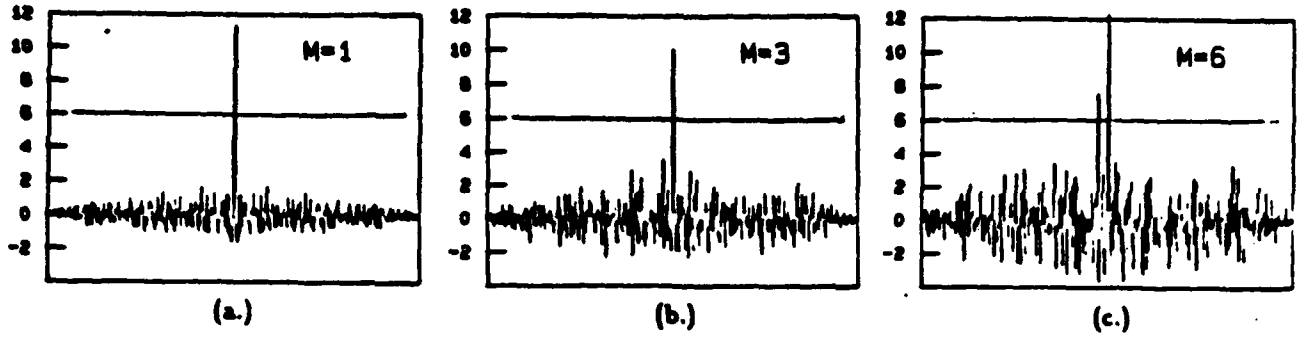


Fig. 2. Digital correlations of a shift invariant filter

In the shift invariant case, the Vander Lugt system performs a correlation between one of the input image $\phi(x, y)$ and the reference filter $h(x, y)$

$$O(x, y) = \sum_{y'=1}^{\sqrt{N}} \sum_{x'=1}^{\sqrt{N}} h(x', y')\phi_i(x' + x, y' + y). \quad [7]$$

For the case where the filter is constructed by simply summing the images in class 1 (multiple exposure) and assuming the same input statistics for each image, the capacity of the shift invariant Vander Lugt system is given by the solution of the following transcendental equation [5]

$$M = \frac{N}{4log(M^3 N)} \quad [8]$$

Asymptotically, the capacity approaches

$$M = \frac{N}{16logN} \quad [9]$$

Thus, the capacity is decreased by only a factor of two from that of the non shift invariant system.

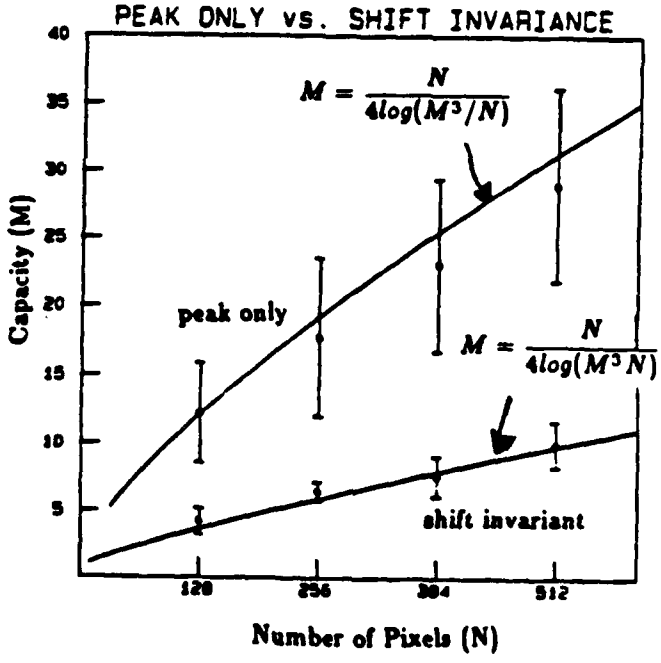


Fig. 3. Capacity curves of the peak only and shift invariant filters

YU 3

This result is important since there is to our knowledge no prior extimate for the loss in capacity due to shift invariance. For the case considered here (the filter derived as a simple sum), the loss is very small; a factor of 2. To verify the theoretical capacity of the correlator, 100 computer trials were averaged to determine the capacity for various N . For each trial, two random vectors were generated to form the initial reference filter. Each image was correlated to determine whether classification was performed correctly. If no error occured, a new random image was added to the reference filter and correlation with all the image was done. The number of images in the reference was increased until a misclassification occured. At this point, the capacity was said to be one less than the number of images stored in the reference.

Figure 3 shows the capacity of both the peak only and shift invariant systems as a function of the number of pixels, N, in the image. Experimental simulations show good agreement with theoretical predictions. It is important to note that because the simulations were performed in the regime of small N, the transcendental equations for the capacity (eqs [5] and [8]) were used to plot the theoretical curves.

IV.Capacity of Binary Filters

As demonstrated above, the capacity of the VanderLugt system can be very large. One potential limitation that might prevent us from actually implementing such a large pattern classification system is the accuracy with which the hologram can record the reference. To get a feel for the susceptibility of the system to nonlinearities and inaccuracies, we considered the capacity of the Vander Lugt correlator when the reference filter has been binarized.

In this case, the reference filter consists of a thresholded version of the filter generated from the multiple exposure algorithm

$$h(x,y) = sgn[\sum_{i=1}^{M} w_i \phi_i(x,y)] \qquad [10]$$

Again, assuming that the input pixels consists of bipolar independent random variables, we find that the capacity of the binary Vander Lugt correlator is asymptotically

$$M = \frac{N}{8\pi log N} \qquad [11]$$

There is only a further $\pi/2$ reduction in capacity from that of the non binarized shift invariant filter.

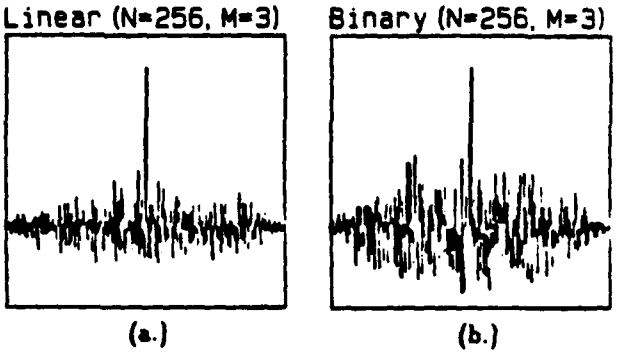


Fig. 4. Digital correlations of the linear and a binary filters

In figure [4], a comparison is made between the linear and a binary filters. In both cases, the input images had 256 pixels and the reference filter contained 3 images. As seen from the figure, the sidelobe level of the binary correlator is significantly larger than that for the linear filter. As a result, as additional images are added to the reference filter, the binary correlator will begin to misclassify sooner. This will correspondingly lead to a lower capacity (theoretical and experimental) for the binary correlator. In figure 5, the capacity of both the binarized and nonbinarized filters are plotted as a function of the number of pixels in the image. Again, computer simulations demonstrate a good agreement with theoretical predictions.

825 22

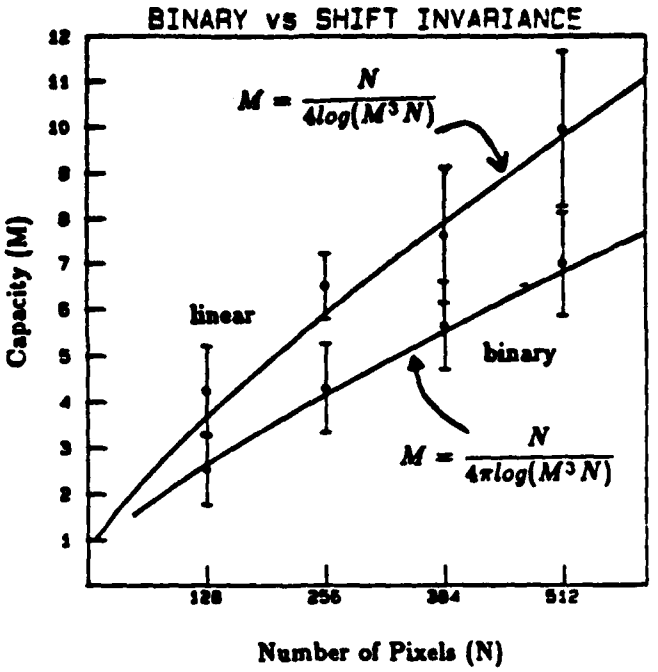


Fig. 5. Capacity curves of a shift invariant and a binary filters.

V.Capacity of the Volume VanderLugt Correlator

In this section, we consider the use of a volume hologram to record the reference filters in a Vander Lugt correlator[6]. We expect that because information is recorded in three dimensions as opposed to the two dimensions for plane holograms, the storage capacity of the volume VanderLugt correlator is increased.

Let us first consider how a volume Vander Lugt correlator operates (Fig.6). Consider the correlation between two point sources. In the recording stage (Fig.6a), the point source generates a plane wave which interferes with a reference wave to form a grating which is recorded in the volume hologram. When an input point source at the same position is presented to the correlator (Fig.6b), a new plane wave reads out the stored grating. The diffracted plane wave is then focussed to form the expected correlation peak at the output. If, however, the input point source is shifted in the direction parallel to the plane of incidence (Fig.6c), the plane wave that is generated will not be Bragg matched with the grating in the volume hologram. Consequently no diffracted wave will be produced and no correlation spot will be formed. In the direction perpendicular to the plane of incidence, the volume hologram exhibits very little Bragg sensitivity and a correlation can still be read out. As a result, shifts of the input in a direction parallel to plane of incidence will not be recognized, while in the perpendicular direction the correlator remains shift invariant.

For an arbitrary input, $A(x,y)$ and reference image, $R(x,y)$, it can be shown that the output of the volume Vander Lugt correlator is [7]

$$O(x,y) = [A(x,y) * R(x,y)]sinc(\alpha x) \qquad [12]$$

where $\alpha = Tsin\theta/2\lambda F$ and $*$ is the correlation operator. T is the thickness of the hologram, θ is the Bragg angle, and F is the focal length of the inverse Fourier transform lens. In other words, the output of the correlator consists of the correlation between the input and reference apodized by a sinc function whose width is determined by the thickness of the volume hologram.

YU 6

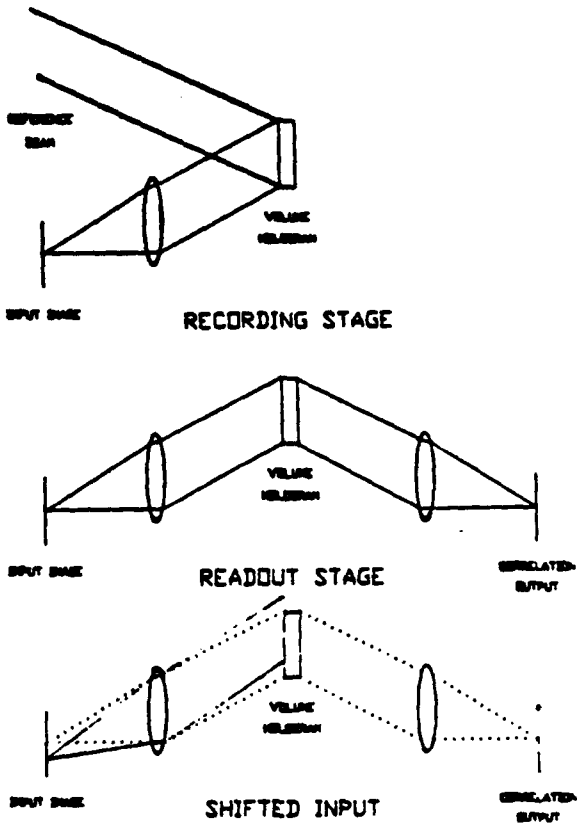


Fig. 6. Recording and readout of a volume hologram.

To experimentally demonstrate this apodising effect, the auto correlation of an O was performed using the volume Vander Lugt correlator. Figure 7 shows a digitally generated autocorrelation of an O which simulates a standard Vander Lugt correlator with a reference O recorded on a plane hologram. In the volume Vander Lugt correlator, the reference O was recorded on a lithium niobate crystal measuring 25x25x5mm. The reference beam was situated such that the plane of incidence was in the horizontal direction. Figure 8a shows the output of the volume Vander Lugt correlator when the input O is positioned at the same plane as the reference O. The output consists of the standard correlation of the two O's multiplied by the horizontal sinc function. When the input O is shifted in the direction parallel to plane of incidence (Fig. 8b), the correlation shifts and only correlation structure to one side of the peak is presented at the output. The smaller spot lying to the right of the primary horizontal band corresponds to the very strong correlation peak lying in the first sidelobe of the apodizing sinc function. Further shifts of the input as shown in figure 8c, merely reads out the correlation structure further from the peak.

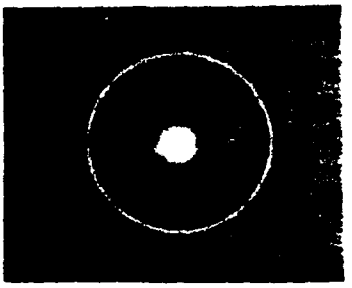


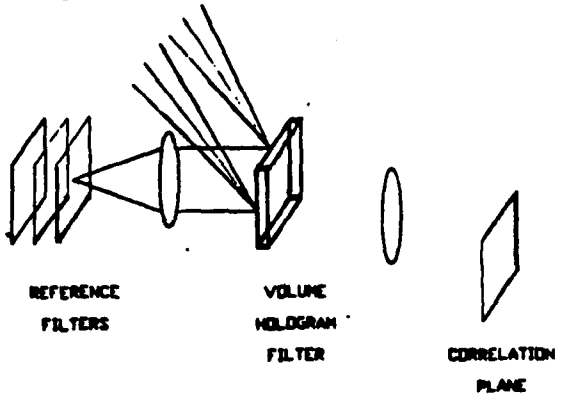
Fig. 7. Digital autocorrelation of an O.

YU 6

(a.) (b.) (c.)

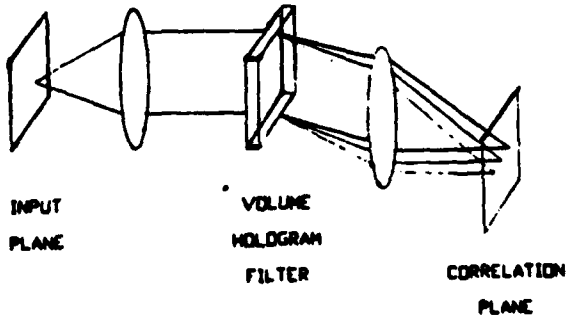
Fig. 8. Experimental outputs of the volume Vander Lugt correlator.

The Bragg selectivity in the volume VanderLugt correlator allows one to perform multi-class categorisation of the input images[8]. In the recording stage (Fig.9a), a set or reference filters is recorded by interfering each with a reference beam separated by the angular bandwidth of the volume hologram. When an input image is presented to the volume Vander Lugt correlator (Fig.9b), a set of correlations is performed simultaneously and presented spatially distributed at the output. The Bragg selectivity of the hologram guarantees that the correlation bands will not interfere with each other. As a result, by detecting which band the correlation peak appear, determines which of many classes the input image belongs to.



RECORDING STAGE

(a.)



READOUT STAGE

(b.)

Fig. 9. Recording and readout stage of the multi-class categorization volume Vander Lugt correlator.

Yu 7

We can consider each correlation band as a seperate output channel performing a simple pattern classification task independent of the other channels. By assuming the same input statistics for the images in each class, the capacity of each output channel can be analytically derived. In this case, the maximum number of images that can be stored was found to be equal to that of the standard VanderLugt correlator [eq.5]. Asymtotically, the capacity of each channel approaches

$$M' = \frac{N}{16 log N} \qquad N \rightarrow \infty \qquad [13]$$

The number of output channels, K, that can be stored in the volume hologram is

$$K = \frac{TL sin\theta}{\lambda F} \qquad (1 \leq K \leq N) \qquad [14]$$

where L is the actual dimension of the output detector array in the direction parallel to the plane of incidence. Hence the total capacity of the system is

$$M_{Total} = \frac{KN}{16 log N} \qquad (N \rightarrow \infty; 1 \leq K \leq N) \qquad [15]$$

Thus the effect of the using a volume hologram is that the capacity is increased by the number of output channels the hologram can support and one to perform multi-class classification. However, one drawback is the partial loss of shift invariance in one direction that results from the use of a volume hologram.

V. Conclusion

In conclusion, we have demonstrated that the capacity of a Vander Lugt correlator without shift invariance is $N/8 log N$ for the simple additive filter. By incorporating the shift invariance inherent in an optical correlator, the capacity is only decreased by a factor 2. Furthermore, by binarising the reference filter, there is a further loss by a factor of $\pi/2$. However, by utilising a volume holgram to record the reference filter, the capacity of the correlator is increased by a factor that can be as high as N with a proportional loss in shift invariance.

Acknowledgements

This research is supported by the Army Research Office and the Air Force Office of Scientific Research. Jeff Yu, is supported by a TRW Program in Advanced Technology Fellowship.

References

[1] A. VanderLugt, *IEEE Trans. Inf. Th.*,IT-10 139 (1964)
[2] D. Casasent *Appl. Opt.*,23 1620 (1984)
[3] A. Lohmann and C. Thum *JOSA*, 23 1503 (1984)
[4] F. Rossnblatt *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan Books, Washington D.C. 1961)
[5] J. Yu, F. Mok, and D. Psaltis *submitted to Applied Optics*
[6] S.I. Stepanov and V. D. Gural'nik *Sov. Tech. Phys. Lett.* ,8 49 (1982)
[7] J. Yu and D. Psaltis *To be published*
[8] C. Guest and C. Gaylord *Appl. Opt.*,19 1201 (1983)

OPTICAL DISK BASED CORRELATION ARCHITECTURES

Demetri Psaltis, Mark A. Neifeld and Alan Yamamura
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

In this paper we describe and experimentally demonstrate optical image correlators that are implemented using optical memory disks. Optical correlation for pattern recognition [1] has long been considered a promising application for optical processing. One of the reasons such correlators have not been used in practical applications yet has been the lack of suitable spatial light modulators to be used as real time input devices. Recently, this limitation has to a large extent been removed through the development of a variety of 2-D SLM's [2] and concepts that allow the utilization of mature 1-D (acoustooptic) SLM's [3]. Attention has therefore shifted to the design of appropriate filters to perform reliable recognition [4]. In most practical applications a single filter is not suffiicient to produce reliable recognition, and the use of spatial [5] and temporal [3] multiplexing to search through a library of filters emerges as the most straightforward solution to the problem. The optical disk correlator architectures we describe in this paper provide an extremely efficient method for performing this task since they combine in a single device the huge memory required for storage of the library of reference images, the spatial light modulator needed to represent the reference in the optical correlator, and the scanning mechanism to temporally search through the library.

The first architecture we will describe is shown in Fig. 1. Each reference image is recorded as a 2-D computer generated Fourier transform hologram on the disk. The input image goes through the beamsplitter, it is Fourier transformed by the lens, and it illuminates the hologram on the disk. The reflected light contains a term proportional to the product of the transforms of the input and reference images. The same lens retransforms the reflected light and the correlation is produced. A principal issue of concern in this architecture is the suitability of commercially available disk systems for recording and reconstruction of holograms. We have identified a write-once disk system which is manufactured with glass (rather than plastic) covers of sufficient optical quality that has allowed us to reconstruct the recorded data using coherent light. We will report the results of this experiment at the conference. The rotation of the disk is used to perform a search through images centered at the same radial position on the disk. An auxillary scanning mechanism is needed in order to position the correlator "head" in the correct radial position. As the disk rotates the entire correlation pattern shifts in one dimension at the output as long as the reference hologram remains in the field of view. A time-delay-and-integrate (TDI) CCD sensor can be used to integrate this traveling correlation pattern in order to improve sensitivity. Alternatively a 1-D parallel read-out detector array can be used that sequentially produces slices of the 2-D correlation pattern as it travels past the detector array.

A straightforward modification of the system of Fig.1 is obtained by recording holograms that are Fourier transforms of the reference images only in the radial dimension since the rotation of the disks provides the necessary shift between the input and reference

along the tracks. The light reflected from such a hologram is Fourier transformed in the radial direction and integrated in the orthogonal dimension onto a 1-D parallel read-out array. The signal from the detector array is again the 2-D correlation presented as a sequence of 1-D slices. The advantage of this architecture compared to the previous one is that it has the same light efficiency as the TDI system without the relative complication of the TDI sensor. Therefore the experiments we will present are with this type of system.

The above architectures require storage of the reference images in the form of computer generated Fourier transform holograms. This provides the advantage of shift invariance which means that we do not need to be concerned with accurate positioning within a single track of the correlation head with respect to the data recorded on the disk. This is a very important practical consideration; the disadvantage however is an increase by a factor of 100 or more in the space bandwidth product required to record the hologram compared to the space bandwidth product of the image itself and an increased computational overhead to record the disk. In addition, the smaller size of the recording results in reduced phase uniformity requirement for the disk. In many cases it is only necessary to record the reference images as binary patterns [6] in which case they can be directly recorded on the disks. Gray scale images can be recorded using some form of area modulation as is done with video disks for example.

There are two types of architecture we will discuss that allow the reference images themselves to be stored on the disk rather than their Fourier transforms. The first is shown in Fig. 2. The input image goes through the beamsplitter and it is Fourier transformed by lens L_1. A Fourier transform hologram of the input is recorded in a photorefractive crystal using a reference beam that is incident from the right, as shown in the figure. Once the hologram is recorded the input is blocked and the the disk is illuminated. L_1 takes the Fourier transform of the reference image that is in the field of view of the illuminating beam and L_2 transforms the light diffracted by the hologram to produce the correlation at the output plane. The rotation of the disk is used to search through a library of images in the radial direction and a TDI detector can be used at the output to increase sensitivity as before. Multiple holograms could be multiplexed in the crystal to address different radial positions on the disk or the entire head can be scanned to address different radial positions as before. We have not yet completed the experimental demonstration of this system but we expect that at the conference we will present the experimental results from this system.

The final architecture we will discuss is shown in Fig. 3. The advantage of this architecture is that it operates on the light intensity and consequently the requirement for phase uniformity is greatly relaxed. As a result it is possible to implement this architecture with most existing disk systems. This correlator works as follows. The reference images are recorded on the disk and the input is imaged through a 1-D scanning device onto the disk. The scanner can be either acoustooptic (as shown in Fig. 3) or a rotating mirror. It provides the relative displacement in the radial direction between the input and reference images that is necessary to calculate the correlation function. The disk rotation provides the displacement in the orthogonal direction. The scanner translates the input image completely accross the stored reference image each time the disk rotates by a distance equal to a pixel of the reference. The intensity of the light reflected from the disk at any one time is proportional to the product between the input and a shifted

version of the reference. The reflected light is collected (integrated) on a single detector which produces as its output a temporal video signal of the 2-D correlation. This system was experimentally demonstrated with acoustooptic scanners. Two types of acoustooptic scanners can be used: A "flying spot" scanner in which a chirp signal propagates in the acoustooptic device acting as a traveling lens that scans the diffracted image at a rate equal to the acoustic velocity. This system completes a scan in a few μs, therefore a complete 2-D correlation takes approximately a few ms. The second scanner that we have demonstrated is a more conventional acoustooptic deflector that scans slowly but permits a higher space-bandwidth product of the input image. A sample of experimental results obtained with the system of Fig.3 is shown in Fig.4. Fig. 4a is a photograph of the pattern recorded on a write-once disk (the acronym CIT) and Fig. 4b is the 2-D correlation produced by the optical system of Fig. 3 and displayed by raster scanning the detector output on a 2-D monitor. Correlations can be produced with our experimental apparatus at rates up to 1000, 100X100 pixel images per second. The optically calculated correlation is in good agreement with the expected autocorrelation function of the CIT pattern. It should be pointed out that since this system operates on intensity we can only represent positive quantities. In order to represent bipolar input and/or reference images we need to add biases at the input stage and subtract it from the output [3], a technique that has been successfully used in a variety of incoherent architectures.

The number of bits that can be stored in the type of disk that we use for most of our work (a write-once, 12 cm diameter system from SONY) is more than 5 billion. The number of 100 × 100-pixel images that can be stored in such a disk is more than 5,000, assuming a generous factor of 100 for loss of spacebandwidth product due to representation (e.g. area modulation for gray scale representation). The rate at which all these images can be interrogated for a possible match with the input is limited by one or more of the following factors: The scanning speed of the disk (40Hz in our case), the speed of the radial scanning mechanism, and the sensitivity and the bandwidth of the output detectors and the electronics following them. As an example consider the system of Fig.2. At 40 Hz disk rotation rate, we obtain 1000 image correlations per 1/40th of a second (i.e. 40,000) image correlations per second), yielding a reasonable 4 MHz bandwidth per detector. It would be extremely difficult to duplicate this capability electronically and it can be achieved with *existing* optical technology. Moreover it is precisely such capability that is required for practical pattern recognition problems.

The research reported in this paper is supported by the Army Research Office.

References

[1] A. Vander Lugt, IEEE Trans. Inf. Theory IT-10, 2 (1964).
[2] OSA Topical Meeting on SLM Technology, Lake Tahoe, June 1988.
[3] D. Psaltis, Opt. Eng. vol 23,1 (1984).
[4] H.J. Caulfield, R. Haimes, and D. Casasent, Opt. Eng. 19(2), 152, (1980).
[5] B.D. Guenther, C.R. Chritensen, and J. Upatnieks, IEEE J.Q.E., vol. QE-15, pp.1348-1362, (1979).
[6] D. Psaltis, Paek, Venkatesh, Opt. Eng. vol. 23, pp.698-704, (1984).

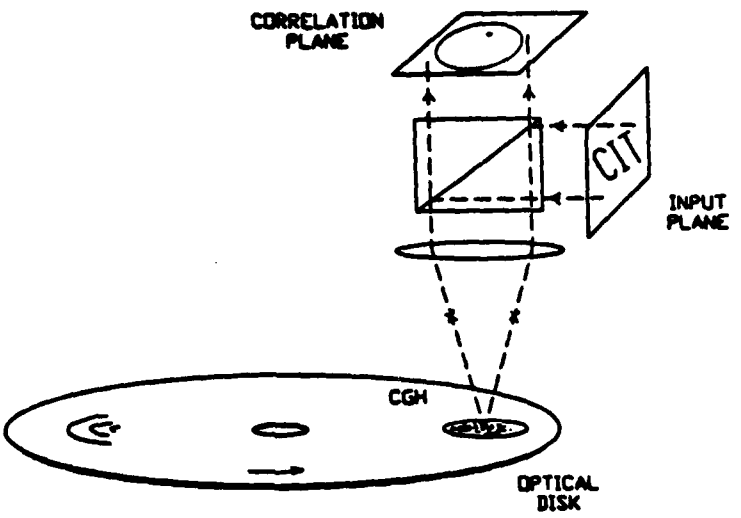


Figure 1

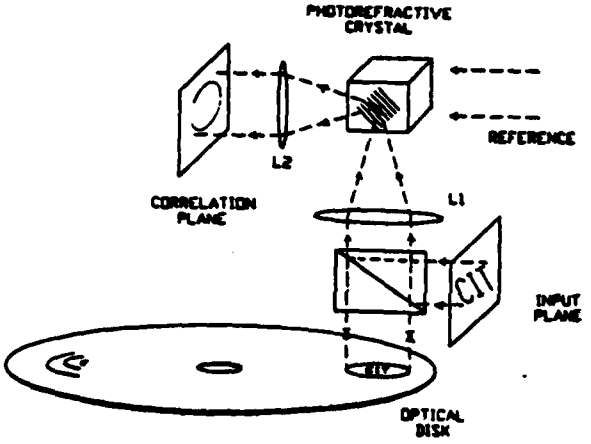


Figure 2

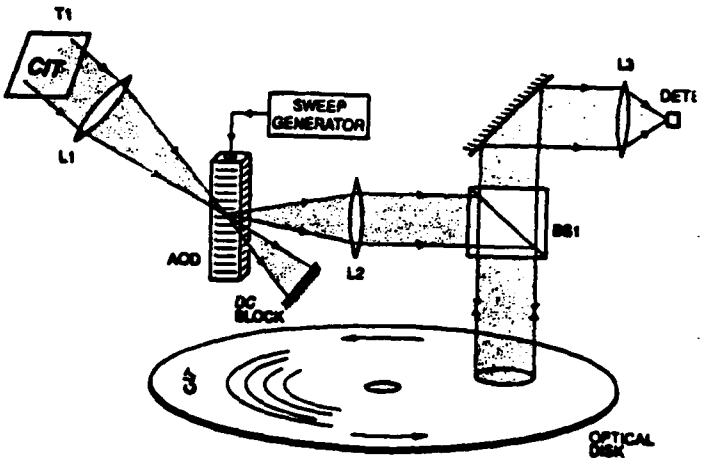


Figure 3

Figure 4a

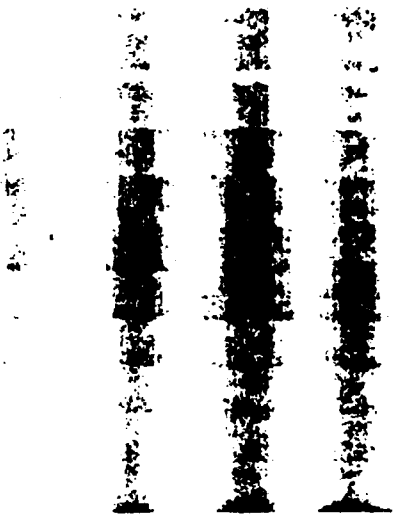


Figure 4b

Optical associative memory using Fourier transform holograms

Eung Gi Paek, MEMBER SPIE
Demetri Psaltis, MEMBER SPIE
California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125

Abstract. An experimental demonstration of a holographic associative memory is presented. The system utilizes an array of classic VanderLugt correlators to implement in parallel the inner product between an input and a set of stored reference images. Each inner product is used to read out an associated image. Theoretical analysis of the system is given, and experimental results are shown.

Subject terms: optical information processing; optical computing; nonlinear optical signal processing; associative memories; neural networks; pattern recognition.

Optical Engineering 26(5), 428-433 (May 1987).

CONTENTS

1. INTRODUCTION

Several types of optical associative memories have been proposed over the years. Ghost image type associative memory was suggested by van Heerden[1] and was investigated by others, including Gabor.[2] In a ghost image holographic associative memory, a hologram of a pattern A is made using the pattern B as the reference. If the hologram is illuminated by A, then the output becomes $A \star A * B$, where $\star$ denotes correlation and $*$ denotes convolution. If A is a noise-like random phase object, then the output can be well approximated by B. If the space-bandwidth product (SBP) of the output images is equal to the SBP of the hologram, then this type of associative memory can store only one pair of associations.[3] If the SBP of the hologram exceeds the SBP of the output patterns, then the number of associations that can be simultaneously stored on the same hologram is equal to the ratio of the two SBPs.[3] Random phase diffusers can be used to improve the quality of the reconstructed images by making the effective bandwidth of the pattern A larger, which makes the $A \star A$ closer to the ideal delta function. The diffuser also makes $A_1 \star A_2$ closer to zero for the cross terms when multiple associations are stored on the same hologram. Willshaw et al.[4] discussed optical memories quite similar to the ghost type and also suggested using thresholding at the output for reducing the cross-correlation "noise" when multiple associations are stored.

A second class of associative memories can be constructed as an array of holographic correlators[5-7] that compare the input and a bank of reference patterns. If a correlation peak is detected above threshold, the associated pattern is produced.

More recently, following the resurgence of interest in neural network models of computation,[8-10] several new holographic memories have been proposed.[11-16] In this paper we are concerned with the holographic memory proposed by Psaltis and Farhat.[11,12] This memory can be thought of as a compromise between the ghost and correlation peak detection memories. We discuss this point with reference to Fig. 1, where the three possible implementations are shown. In the ghost image memory shown in Fig. 1(a), the input pattern $\hat{f}$ is correlated against all stored memories, and the correlation function is then convolved with the associated stored output pattern. The final result is obtained by summing all of the reconstructed images. The peak detection memory shown in Fig. 1(b) detects the presence of the peak in the correlation plane to determine the best match between the input and one of the stored input images. Once the match has been established, only the corresponding memory is illuminated, which eliminates the crosstalk and the distortion present in the ghost holograms. The final possibility, shown in Fig. 1(c), is to sample the correlation plane only at the origin, where autocorrelation peaks occur, using an array of pinholes (rather than to actually detect the peak). The spatially sampled correlation peak rather than the entire correlation plane is then convolved with the associated stored output image. This eliminates the distortion present in the ghost image holograms but does not entirely eliminate the crosstalk. Consequently, compared to the ghost holograms, the quality of the recalled images is dramatically improved in this case; however, the number of patterns that can be stored in the same hologram is reduced, compared with the peak detection type. The crosstalk at the output of the system can be reduced by thresholding if the stored patterns are binary, and further improvement can be realized for autoassociations (i.e., the input and output stored patterns are the same) through the use of feedback.[8-10] The advantages of the latter type of memory are its simplicity, since no active devices are needed at the intermediate level, and its robustness with respect to failure of components. In the peak detection memory, if the element that senses the correlation peak of a particular stored pattern fails, the entire memory is erased. Moreover, this type of memory (also referred to as outer product memory) is widely used in the modeling of neural networks, and it generalizes very naturally to multilayered networks.

In this paper we describe in detail two holographic associative memories that utilize the pinhole sampling method. Both memories are constructed as an array of VanderLugt correlators. Multiple images are multiplexed on Fourier transform holograms. A single hologram is used in the first architecture.

Invited Paper IP-107 received Jan. 29, 1987; revised manuscript received March 13, 1987; accepted for publication March 13, 1987; received by Managing Editor March 13, 1987.

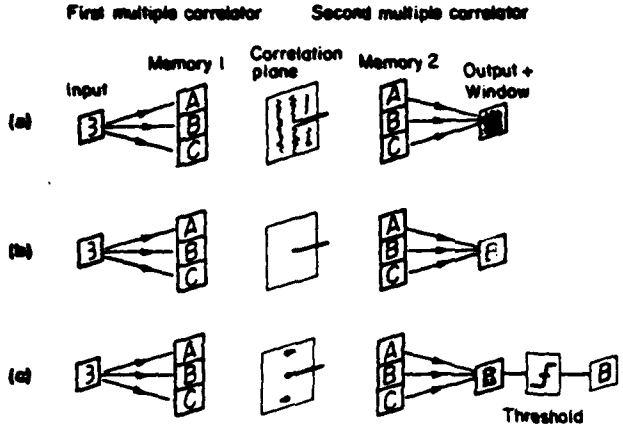


Fig. 1. Classification of optical associative memories. (a) Ghost image; (b) peak detection; (c) pinhole sampling.

an autoassociative memory, since the input and output patterns are the same. The second architecture utilizes two separate holograms for storing the input and output patterns and can be either auto- or heteroassociative. The experimental demonstration of both systems operating as autoassociative memories is described.

2. ARCHITECTURES

To construct a two-dimensional outer product associative memory, we need to implement the following operation:

$$\dot{h}(x,y) = \int T(x,y,\xi,\eta)\dot{f}(\xi,\eta)d\xi\,d\eta\,, \tag{1}$$

where

$$T(x,y,\xi,\eta) = \sum_{m=1}^{M} h_m(x,y)f_m(\xi,\eta)\,. \tag{2}$$

$T(x,y,\xi,\eta)$ is the synaptic matrix in accordance with the outer product storage mechanism, $f_m(\xi,\eta)$ is the *mth* input memory, $h_m(x,y)$ is the associated output pattern, and M is the number of memories stored. $\dot{f}(\xi,\eta)$ is the input function to the memory, and $\dot{h}(x,y)$ is its output. In the remainder of this paper we concern ourselves with autoassociative memories, in which $h_m(x,y) = f_m(x,y)$. To implement Eq. (1), we need a four-dimensional interconnection matrix that cannot be directly implemented optically. However, if we substitute Eq. (2) into Eq. (1) and rearrange, we end up with the following inner product representation:

$$\dot{h}(x,y) = \iint\left[\sum_{m=1}^{M} f_m(x,y)f_m(\xi,\eta)\right]\dot{f}(\xi,\eta)d\xi\,d\eta$$

$$= \sum_{m=1}^{M}\left[\iint f_m(\xi,\eta)\dot{f}(\xi,\eta)d\xi\,d\eta\right]f_m(x,y)\,. \tag{3}$$

From Eq. (3) we deduce that the optical implementation of this memory can be decomposed into three steps. First, the inner product of the input and each memory must be formed. This can be optically evaluated as the correlation sampled at the origin. Second, each inner product must be multiplied by the associated memory. Finally, these products must be summed over all of the memories to produce the final result.

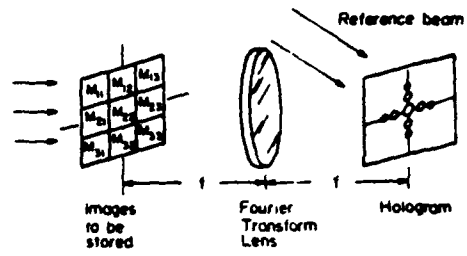


Fig. 2. Optical setup for the recording of the holographic memory.

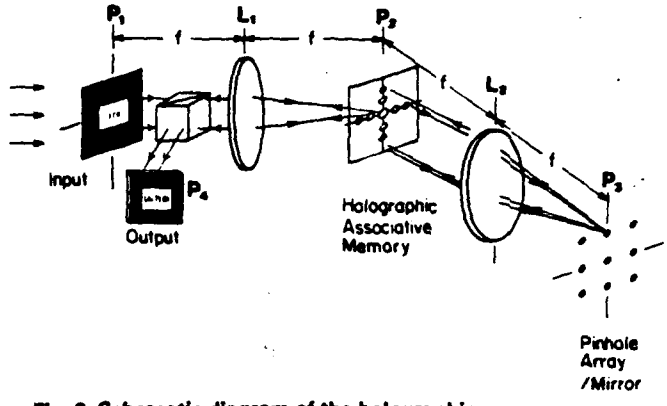


Fig. 3. Schematic diagram of the holographic memory system.

The images are stored in a conventional Fourier transform hologram, as shown in Fig. 2. All of the memories to be stored are arranged side by side, spatially separated from each other. In the analysis, it is assumed that each memory is separated by the same distance along the x and y directions. The Fourier spectrum of all of the memories interferes with a single tilted plane wave to simultaneously make a multiple hologram. The amplitude distribution at the input plane is

$$\sum_{m=1}^{M} f_m(\xi - a_m, \eta - b_m)\,, \tag{4}$$

where a_m and b_m are the positions of the *mth* image in the ξ and η directions, respectively. When we record the interference on a holographic plate at the hologram plane, the amplitude transmittance of the hologram becomes

$$\left|\sum_{m=1}^{M} F_m(u,v)\exp\left[-j(ua_m + vb_m)\right] + \exp(-ju\xi_0)\right|^2$$

$$= \sum_{m=1}^{M} F_m^*(u,v)\exp\{j[u(a_m - \xi_0) + vb_m]\}$$

$$+ \text{conjugate term} + \text{dc terms}\,, \tag{5}$$

where $F_m(u,v)$ is the Fourier transform of $f_m(x,y)$ and ξ_0 is a constant that determines the angle of incidence of the reference beam.

The system used to recall the information stored on the hologram, shown in Fig. 3, is a modified VanderLugt correlator. The input pattern is placed at plane P_1 and is Fourier

transformed by lens L_1. The transform illuminates the holographic memory placed at the Fourier plane P_2. The correlations of the input and each memory are produced at plane P_3 by lens L_2. The inner product values are sampled by an array of pinholes at P_3. Each pinhole is positioned exactly where each of the stored images was centered when the Fourier transform hologram was recorded. Therefore, if the input is one of the stored images centered on the optical axis, then a sharp autocorrelation peak will form at P_3 on one of the pinholes. Light emerging from each pinhole is reflected by a mirror placed immediately after the pinholes, and the reflected light illuminates the hologram to form the reconstructed images of all of the memories at the output plane P_4. The reconstruction due to light from each pinhole is the entire composite memory shifted by the position of the pinhole. At the origin at plane P_4 we obtain the superposition of all of the memories. The strength with which each memory is represented in this superposition is proportional to the value of the inner product between the input and the corresponding memory. A window is placed at P_4 to select only the desired central portion of the reconstructed holographic images.

We now describe the operation of this system analytically. Let $\hat{f}(\xi, \eta)$ be the amplitude transmittance at plane P_1 in Fig. 3. Then, the term of interest in the amplitude of the light diffracted by the hologram is

$$\sum_{m=1}^{M} \hat{F}(u,v)F_m^*(u,v)\exp\{j[u(a_m - \xi_0) + vb_m]\} . \tag{6}$$

At the correlation output plane P_3, the light amplitude is the Fourier transform of Eq. (6):

$$\sum_{m=1}^{M} g_m(-x', -y') * \delta\{x' - a_m + \xi_0, y' - b_m\} , \tag{7}$$

where $g_m(x',y')$ is the correlation of $f_m(\xi,\eta)$ with $\hat{f}(\xi,\eta)$, and x',y' are the coordinates in plane P_3. The correlation output is sampled by the pinhole array located at coordinates $x' = a_m - \xi_0$, $y' = b_m$ in P_3. We assume that the pinholes can be adequately described mathematically as delta functions. Then, the light reflected by the mirror at plane P_3 can be written as

$$\sum_{m=1}^{M} \left[g_m(-x',-y') * \delta\{x' - a_m + \xi_0, y' - b_m\} \right] \times \delta\{x' - a_m + \xi_0, y' - b_m\}$$

$$= \sum_{m=1}^{M} g_m(0,0)\delta\{x' - a_m + \xi_0, y' - b_m\} . \tag{8}$$

The reflected light illuminates the hologram in P_2, and the amplitude of the light traveling from right to left in Fig. 3 immediately to the left of P_2 is given by

$$\sum_{m=1}^{M}\sum_{m'=1}^{M} g_m(0,0)F_{m'}^*(u,v)\exp\{-j[u(a_m - a_{m'}) + v(b_m - b_{m'})]\} . \tag{9}$$

The light at the output plane P_4 is the Fourier transform of

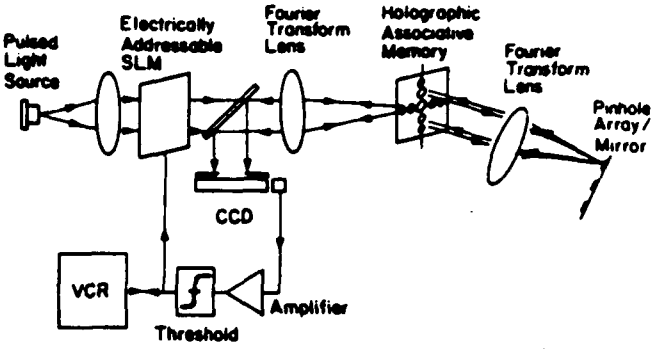


Fig. 4. Closed-loop version of the holographic memory system.

Eq. (9). Note that in the above equation, unless $m = m'$, the spectra $F_m(u,v)$ will emerge on a high spatial frequency carrier, which means that they will be reconstructed off-axis at P_4. The total light amplitude at P_4 can be written as

$$\sum_{m=1}^{M}\sum_{m'=1}^{M} g_m(0,0)f_{m'}^*(x + a_m - a_{m'}, y + b_m - b_{m'}) . \tag{10}$$

When we observe the light only through a window that is centered around the optical axis and is equal in size with each memory, only the terms $m = m'$ survive:

$$\sum_{m=1}^{M} g_m(0,0)f_m^*(x,y) . \tag{11}$$

If f_m is real,

$$\sum_{m=1}^{M} g_m(0,0)f_m(x,y) . \tag{12}$$

Comparing the result in Eq. (12) with Eq. (3), we see that the optical system we described is exactly the outer product associative memory.

If the input pattern is most similar to the stored image $f_{m_0}(\xi,\eta)$, then the correlation between the input and f_{m_0} will be the strongest, and consequently [from Eq. (12)], $f_{m_0}(x,y)$ will be amplified the most in the final output reconstruction. However, there is still crosstalk, since all of the other memories are also weakly read out. This crosstalk can be eliminated if the stored images are binary, in which case the output can be thresholded and fed back to the input for multiple iterations.[8] A closed-loop version of the holographic memory is shown in Fig. 4. The light at the output is detected on a two-dimensional CCD. The video signal from the CCD is electronically thresholded and fed back to the input plane of the system through an electronically addressed spatial light modulator. The Litton magneto-optic spatial light modulator[17] is one candidate device that can be used for this purpose. The system also can be configured with optical feedback using an optically addressed spatial light modulator.

An alternative implementation of this type of associative memory is shown in Fig. 5. This architecture is basically an unfolded version of the system in Fig. 3. In other words, instead of having a mirror that reflects the light back through the same system, we have two identical optical systems, one after the other. This allows us to use two separate holograms.

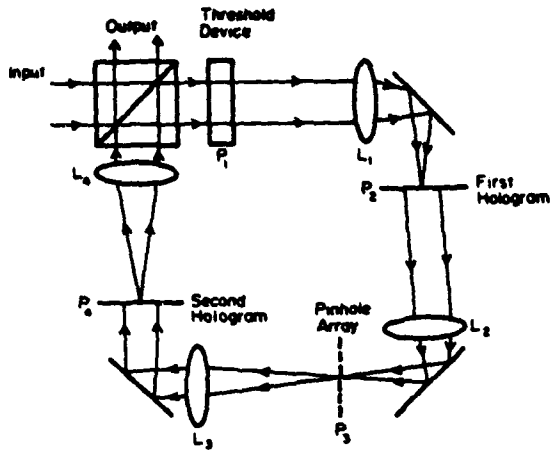


Fig. 5. Optical associative loop.

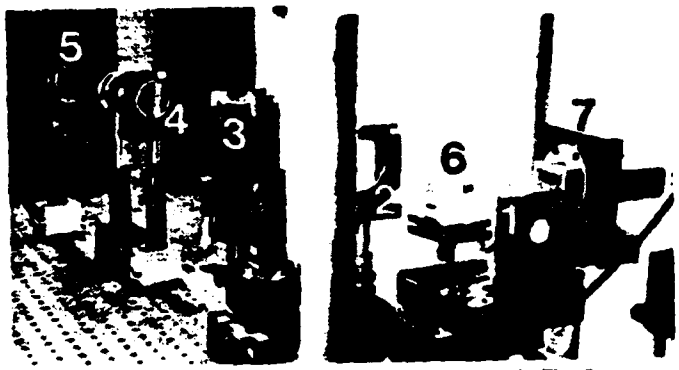


Fig. 6. Experimental setup of the system in Fig. 3.

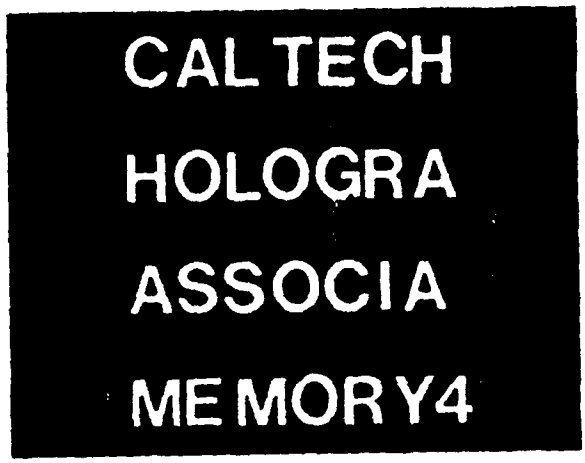
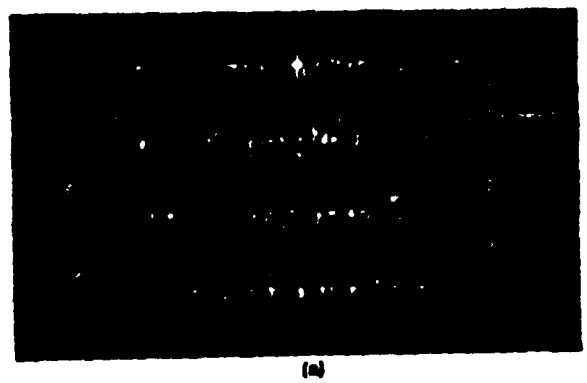


Fig. 7. The four patterns stored in the holographic memory for the experiment.



(a)

(b)

Fig. 8. Correlation outputs for (a) the first and (b) the third memories at the input.

which provides added flexibility in designing the system and also makes heteroassociations possible.[18,19] The system shown in Fig. 5 is an autoassociative memory with feedback and thresholding. An input image enters the system through the beamsplitter, as shown in the figure, and is thresholded by an optically addressed spatial light modulator. Several devices can be used for this purpose, such as the Hamamatsu microchannel spatial light modulator[20] or the Hughes liquid crystal light valve.[21] The optical system from plane P_1 to plane P_3 is a modified VanderLugt correlator similar to the one used previously. The correlation patterns are sampled by an array of pinholes at P_3, and the light emerging from plane P_3 illuminates a second, identical system. The light reaching plane P_1 is the superposition of all of the images that have been stored in the multiplexed holograms. Each image is weighted by the inner product between the pattern recorded on the spatial light modulator from the previous iteration and itself. Thus, the systems shown in Figs. 4 and 5 are functionally identical. As will be seen when we describe the experimental demonstration of the two systems, the added flexibility of the system in Fig. 5 can significantly improve the performance.

3. EXPERIMENTAL RESULTS

The experimental apparatus assembled to demonstrate the memory of Fig. 3 is shown in Fig. 6. The multiplexed Fourier transform hologram (item 3) was fabricated in dichromated gelatin. The pinhole array (5) was made by drilling four holes on a thin metal plate. The diameter of each pinhole was 350 μm. The pinhole array was then placed in contact with a mirror. A CCD camera (7) was used to detect the output of the memory through the beamsplitter (6). The four patterns used as the memories in this experiment are shown in Fig. 7. The patterns obtained at the correlation plane (or equivalently, the plane of the pinholes) when the first and third stored patterns were presented at the input are shown in Figs. 8(a) and 8(b), respectively. A sharp autocorrelation peak is evident in both cases, and the position of these peaks coincides precisely with the position of two of the four pinholes. It is interesting to note that the inclusion of the pinholes destroys the shift invariance of the VanderLugt correlator. If the input pattern is

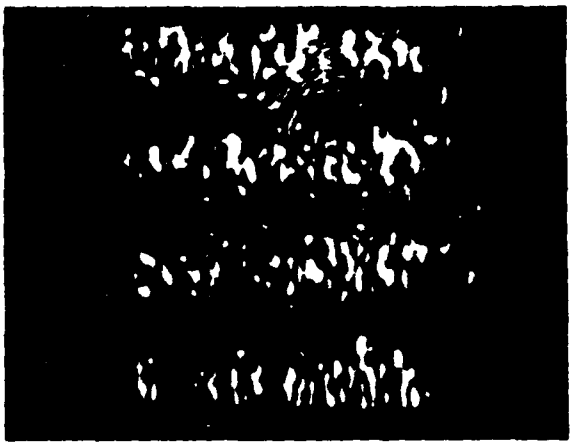


Fig. 9. Recalled image produced when the pinholes were removed and the first memory (CALTECH) was placed at the input plane.

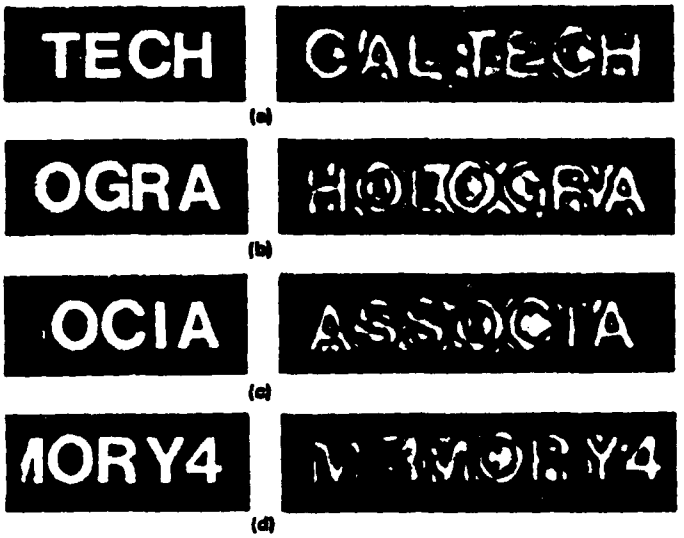


Fig. 10. Partial inputs (left) and the associative recalled outputs (right).

shifted from its nominal position, then the correlation peak shifts as well; consequently, it misses the pinhole, and thus no light is reflected. If the pinholes are removed, however, we are back to ghost holography. The recalled image obtained when the pinholes are removed and the first memory is placed at the input plane is shown in Fig. 9. This image was obtained at the output of the system (on the CCD) with the masking window removed. The image is obviously highly distorted, and there is no apparent favoring of the correct memory. A dramatic improvement is obtained when the pinholes are included (Fig. 10, with partial inputs also displayed). In this case, each memory is faithfully reconstructed. The shift invariance of the system can be restored by including a quadratic nonlinearity in the correlation plane.[22,23]

The experimental apparatus of the loop system (Fig. 5) is shown in Fig. 11. The system was constructed with the microchannel spatial light modulator as the threshold device, and the holograms were recorded on thermoplastic plates. The four Fourier transform lenses and the pinhole array are visible in the photograph. Several results obtained with this setup are shown in Fig. 12. The four faces used as the memories are displayed in Fig. 12(a). Figures 12(b) and 12(c) show the

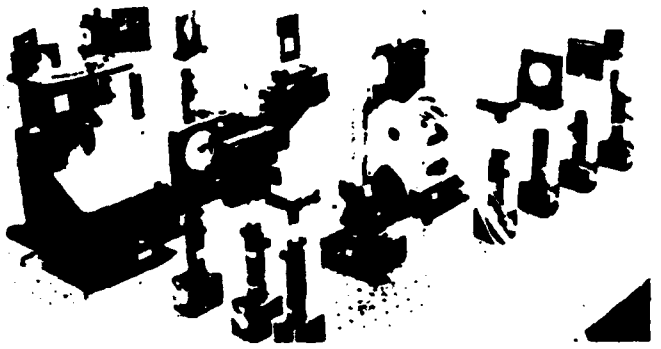


Fig. 11. Experimental setup of the loop.

reconstruction of the first and second holograms, respectively (see Fig. 5). Note that the reconstruction of the first hologram, used for recognition in this architecture, is edge enhanced. This was accomplished by recording the hologram such that the high spatial frequency portion of the spectrum was enhanced, which ensures that the cross-correlations between the four faces are much smaller than the autocorrelation peaks. The second hologram, used for read out of the stored information, is recorded so that a faithful reconstruction is obtained using diffuse illumination during the recording. The partial input and the complete recalled image are shown in Figs. 12(d) and 12(e), respectively. The noise evident in Fig. 12(e) is speckle, a consequence of the diffuser used to form the second hologram. Comparing Fig. 12(e) with Fig. 10, we see that in Fig. 10 there is still evidence of crosstalk superimposed on the reconstructed images, while no crosstalk is detectable in Fig. 12(e). The thresholding performed by the spatial light modulator, the high space-bandwidth product of the images used, and the virtual orthogonalization of the four memories accomplished by the high pass filtering in the first stage combine to eliminate the crosstalk in a single pass through the loop in the second experiment.

4. ACKNOWLEDGMENTS

We would like to thank N. Farhat and K. Hsu for many helpful discussions. We also thank Hamamatsu for donating to us the spatial light modulator used in some of the experiments. This research is supported by the Defense Advanced Research Projects Agency, the Army Research Office, and the Air Force Office of Scientific Research.

5. REFERENCES

1. P. J. van Heerden, "A new optical method of storing and retrieving information," Appl. Opt. 2, 387–392 (1963).
2. D. Gabor, "Associative holographic memories," IBM J. Res. Dev. 13, 156–159 (1969).
3. J. Hong and D. Psaltis, "Storage capacity of the holographic associative memories," Opt. Lett. 11(12), 812–814 (1986).
4. D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, "Non-holographic associative memory," Nature 222, 960–962 (1969).
5. M. Sakaguchi, N. Nishida, and T. Nemoto, "A new associative memory system utilizing holography," IEEE Trans. Comp. C-19, 1174–1181 (1970).
6. G. R. Knight, "Page-oriented associative holographic memory," Appl. Opt. 13, 904–912 (1974).
7. G. C. Guest and T. K. Gaylord, "Truth-table look-up optical processing utilizing binary and residue arithmetic," Appl. Opt. 19, 1201–1207 (1974).
8. J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. Natl. Acad. Sci. USA 79, 2554–2558 (1982).
9. T. Kohonen, *Self-Organization and Associative Memory*, Springer-

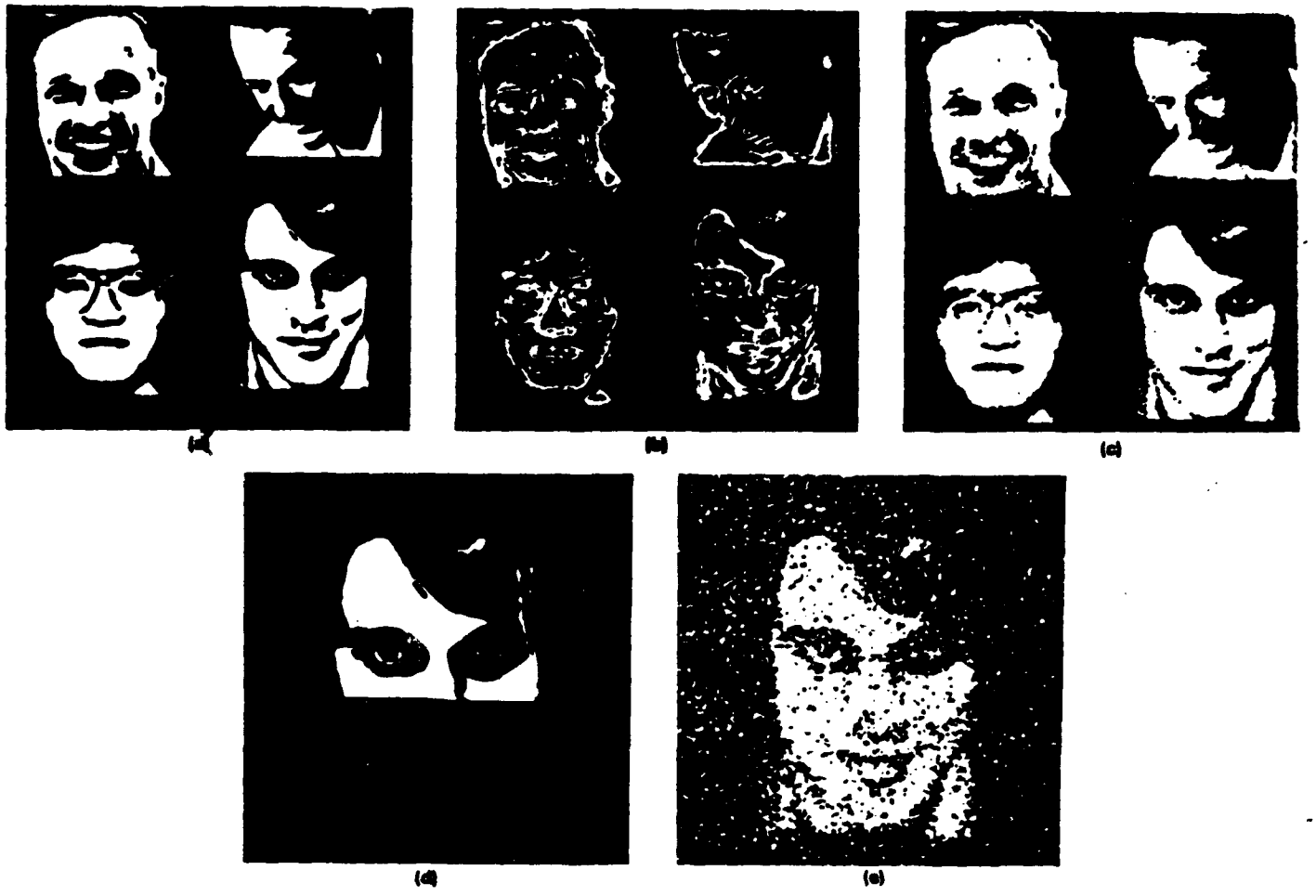


Fig. 12. (a) The four memories stored; the reconstructed images from (b) the first hologram and (c) the second hologram; and (d) partial input and (e) the recalled output.

Verlag, New York (1984).

10. J. A. Anderson, "A simple neural network generating an interactive memory," Math. Biosci. 14, 197–220 (1972).
11. D. Psaltis and N. Farhat, "New approach to optical information processing based on the Hopfield model," in *Tech. Digest ICO-13 Conf.* (Sapporo, Japan), pp. 24–25 (1984).
12. D. Psaltis and N. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," Opt. Lett. 10(2), 98–100 (1985).
13. B. H. Soffer, G. H. Dunning, Y. Owechiko, and E. Marom, "Associative holographic memory with feedback using phase-conjugating mirrors," Opt. Lett. 11, 118–120 (1986).
14. D. Anderson, "Coherent optical eigenstate memory," Opt. Lett. 11(1), 56–58 (1986).
15. A. Yariv and S. K. Kwong, "Associative memories based on message-bearing optical modes in phase conjugate resonators," Opt. Lett. 11(3), 186–188 (1986).
16. H. - K. Liu, S. Y. Kung, and J. A. Davis, "Real-time optical associative retrieval technique," Opt. Eng. 25(7), 853–856 (1986).
17. W. E. Ross, D. Psaltis, and R. H. Anderson, "Two-dimensional magneto-optic spatial light modulator for signal processing," Opt. Eng. 22(4), 485–490 (1983).
18. D. Psaltis and E. G. Paek, "Holographic heteroassociative memory," to be submitted to Appl. Opt.
19. E. G. Paek and D. Psaltis, "Holographic implementation of a neural network model," J. Opt. Soc. Am. A3, 17(1986).
20. C. Warde and J. Thackara, "Operating modes of the microchannel spatial light modulator," Opt. Eng. 22(6), 695–703 (1983).
21. W. P. Bleha et al., "Application of the liquid crystal light valve to real-time optical data processing," Opt. Eng. 17(4), 371–384 (1978).
22. D. Psaltis and C. H. Park, "Nonlinear discriminant functions and associative memories," in *Neural Networks for Computing*, J. Denker, ed., AIP, New York (1986).
23. D. Psaltis and J. Hong, "Shift-invariant optical associative memories," Opt. Eng. 26(1), 10–15 (1987).

Eung Gi Paek received the BS degree in physics from Seoul National University in 1972 and the MS and Ph.D. degrees in physics from Korea Advanced Institute of Science and Technology in 1976 and 1979 respectively. From 1979 to 1982 he worked at the Agency for Defense Development in Korea. In 1982 he joined the California Institute of Technology and currently is a senior research fellow working in the areas of hybrid optical signal/image processing, pattern recognition, and optical implementations of neural network models.

Demetri Psaltis received B.Sc. degrees in electrical engineering and economics in 1974 and the M.Sc. and Ph.D. degrees in electrical engineering in 1975 and 1977, respectively, all from Carnegie Mellon University. After completing his Ph.D. dissertation, he remained at Carnegie Mellon as a research associate and later as a visiting assistant professor for a period of three years. In 1980 he joined the faculty of the Electrical Engineering Department at California Institute of Technology, where he is now an associate professor and a consultant to industry. His research interests are in the areas of optical information processing, image processing, pattern recognition, neural network modeling, and optical devices, and he has published over 100 technical papers in these areas. Dr. Psaltis is a fellow of the OSA and a member of SPIE.

To appear in the Proceedings of the IEEE Annual International Conference on Neural Networks, San Diego, July 24-27, 1988.

INVARIANCE AND DISCRIMINATION PROPERTIES OF THE OPTICAL ASSOCIATIVE LOOP

Ken Hsu and Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125

Introduction

In this paper we report recent experimental results from the optical associative memory that we have described previously [1,2]. This system is a single layer neural network architecture simulating a 2-D array of approximately 10^5 neurons on which images can be represented. This 2-D array of neurons is fully interconnected via holograms and the system is organized as an auto-associative memory with feedback. An external image projected into the system causes one of the stored images to become a stable state of the system. The ability of the system to recognize distorted versions (e.g. rotated, shifted, or scaled) of a stored image depends critically on the gain of the system as the light goes around the loop. High gain provides invariance to distortions but ultimately it also leads to a loss in discrimination against unfamiliar images. Thus there is an optimum choice of parameters of the system that yields optimum performance. In what follows we describe how the parameters affect the performance of the memory and we report the performance (in terms of discrimination vs. invariance) obtained by the experimental system.

Experimental System

A schematic diagram of the optical associative loop is shown in Fig.1 and a photograph of the experimental apparatus is shown in Fig.2. This processor is comprised of two cascaded correlators of which the first is used for calculating the degree of similarity between the external input image and the images stored in the hologram. The second correlator uses the output from the first correlator to reconstruct the same images that are also stored in the second hologram to provide the feedback signal for the loop. The operation of this associative loop can be explained with the aid of the block diagram shown in Fig.3a. In this example four images spatially separated and stored in the Fourier transform holograms H_1 and H_2 as shown in Fig.3b. When the input pattern **A** is presented as an input to the system, the first correlator produces the auto-correlation pattern along with three cross-correlations at plane P_2. The pinhole array at P_2 samples these correlation patterns at the middle of each pattern where the inner products between the input and each of the stored images form. Each of the four beams that go through the pinholes goes through the second correlator to reconstruct the four images stored in hologram H_2. These reconstructed images are spatially separated and superimposed at plane P_1. The stored image which is most similar to the input pattern gives the strongest correlation signal hence the brightest reconstructed image. The weakly read-out from the cross-correlation can be eliminated by thresholding by the LCLV. The output of the LCLV becomes the new input image for the loop and thus iterations take place. The stable pattern that forms as a recirculating image in the loop is the stored image that is most similar to the original

input.

In the system of Fig.1 the input pattern is imaged onto the LCLV by lens L_i and through beam splitter BS_3. A collimated argon laser beam illuminates the read-out side of the LCLV through beam splitters BS_2 and BS_1. A portion of the reflected light from the LCLV that propagates straight through BS_1, is diverted by BS_2, and it is imaged by lens L_0 onto a CCD television camera. This provides real time monitoring of the activity of the system. The portion of the light that is reflected by BS_1 into the loop is Fourier transformed by lens L_1 and illuminates the hologram H_1. The correlation between the input image and each of the stored images is displayed at plane P_2. The pinhole array at P_2 has center spacings corresponding to the spatial separations of the stored images. The remainder of the optical system from P_2 back to the neural plane P_1 is essentially a replica of the first half, with the hologram H_2 storing the same set of images at H_1. Fig 4 shows an example of an experiment performed with this loop. Fig. 4a is the external input, in this case a partially obstructed image of one of the stored patterns. Fig.4b shows the response of the system with the external input still present, and Fig.4c shows the stable state of the loop after the external input is removed.

One of the interesting properties of this system is its dynamics. The time for the loop to reach a stable state depends very much on the initial conditions. Fig.5a shows the temporal response of the loop to an input pattern. When the signal in the lower trace becomes high, it indicates that the external input is ON. The upper trace shows the corresponding response of the loop. The initial rise is due to the presentation of the input whereas the second rise is due to the fact that the feedback path was closed. It is seen that it takes about two seconds for the loop to reach a stable state whereas the rise time of the LCLV is approximately one second in the mode we operated it. When the external input is turned off, the loop remains latched to a stable state which is one of the stored images. Fig.5b shows the same experiment but with input intensity reduced to one third of the first input. The second rise of the upper trace shows that it takes approximately four seconds for the loop to reach its stable state. After the input is turned off the loop gives the same output intensity. This example shows that initial conditions affect the dynamics of the loop but it does not affect its final state. We will see in the next section similar invariances when the input is shifted and rotated.

The loop dynamics and related invariance properties can be best understood by using a network model as shown in Fig.6. Each resolution element of the LCLV simulates a separate neuron and with resolution of the device used being approximately 400 × 400 pixels, 160,000 neurons are simulated. Each of these neurons is globally connected and fed back to everyone via the two holograms. The optical signal is attenuated in the loop due to the diffraction efficiencies of the Fourier transform holograms and the losses from pinholes as well as lenses and beam splitters. Therefore neurons have to provide optical gain to compensate this loss. In our system this is achieved by adding an image intensifier at the photoconductor side of the LCLV. The microchannel plate of the image intensifier is sensitive to a minimum incident intensity of approximately 1 nW/cm^2 and it reproduces the input with an intensity 10^4 times brighter ($10\ \mu W/cm^2$). This is bright enough to drive the LCLV. If we use a beam with intensity equal to 10 mW/cm^2 to read the LCLV then the intensity of the output light is approximately 1 mW/cm^2. Thus, the combination of the image intensifier and the LCLV provide optical gain up to 10^6. Fig.7a shows the input-

output characteristics of the optical thresholding element which is similar to a sigmoid function. The optical gain can be adjusted by changing the bias voltage of the image intensifier. Fig.7b shows the relationship between the bias voltage applied to the image intensifier and the gain.

The dynamics of the recall process can be described by using an iteration map formed by the gain and loss curves as shown in Fig.8. In the figure the slope of the straight line is proportional to loop loss due to the holograms and the pinholes and it is superimposed with the input-output response of the neurons. The intersection points of this line with the neuron gain curve at point Q_1 determines the threshold level and Q_2 represents a stable point. If the initial condition of the neuron is above the threshold point $\theta 1$, the signal grows in each iteration until it arrives and latches at Q_2. On the other hand, if the initial condition is below $\theta 1$ the signal will decay to zero. The number of iterations depends on the distance of the initial condition from the threshold. This explains the dynamics of Fig.5. Similarly, if the loop loss is low or the neuron gain is high one can expect that the loop will converge faster to a stable state. Raising the gain also has the effect of lowering the threshold of the system. In the following section we will see that the setting of the gain is the key parameter that mediates the trade-off between distortion invariance and discrimination capability of the loop.

Invariance versus Discrimination Trade-offs

In the previous section we saw that as long as the gain is high enough and the external input is strong enough to produce an initial condition for the LCLV that is above threshold, then the loop will converge to one of the stored stable states. Since the external input does not affect the shape of the final state, but rather it selects which state is produced we can build a degree of invariance in the system since a shifted, rotated or scaled version of a stored image can cause the stored image to be recalled. The effect of such distortions of the input image are to decrease the level of the initial condition. Therefore, by raising the neuron gain, no matter how much we change the initial condition by rotating, shifting, and scaling the input image, the loop can always be made to produce an image as a stable state. But the ability to correctly recognize a stored image from a distorted input and the discrimination capability, i.e. the ability to distinguish images from one another are two things that compete with each other. If there is too much gain then just shining a flush light at the input of the system causes it to lock on to one of its stable states. If the gain is set too low then even an input that is a slightly distorted version of one of the stored images is not recognized. There are two parameters under our control that can affect the gain in the loop: The gain of the neurons and the size of the pinhole.

We will use Fig.3b as an example. Let $f_i(x,y), i = 1,2,3,4$, represent the images of the letters A, B, C, D, respectively and let the pinhole size be W. Then the reconstructed images in the window at P_1 can be shown to be

$$\sum_{i=1}^{4}[g_{1i}(x,y)rect(\frac{x}{W})rect(\frac{y}{W})] * f_i(x,y)$$

where * represents the convolution operation, $g_{11}(x,y)$ is the auto-correlation of A and $g_{1i}, i \neq 1$, are the cross-correlations of A with B, C, D, respsctively. We see that the images are blurred by the finite dimension of the pinholes. Decreasing W gives better

image quality but we need to increase the gain of the neurons to compensate for the loss due to the small pinholes. At the other limit, if the pinhole size is increased we do not need very high gain neurons but the image quality deteriorates. In the limit when W becomes infinitely large, the reconstructed image in the window at P_1 becomes a superposition of all the stored images, each equally strong, and severely blurred. Thus there is an optimum pinhole size and an optimum neuron gain. Fig.9 shows the minimum gain required and maximum gain allowable for the loop to sustain a stable memory as a function of pinhole size. Below the minimum gain the loop can not recognize any image in the sense that once the external input is cut off the loop activity decays to zero. Above the maximum gain the loop loses discrimination capability meaning that any input image even a flash light will induce the loop into a stable state. This behavior is consistent with our previous predictions. Note that the minimum gain increases when the pinhole size is increased to more than 250 μm. This is because the reconstructed images are blurred so much that the correlation peaks are weakened and the losses in the loop are increased. Fig.9 shows that the optimum pinhole size in this system is in the range of 70 μm to 150 μm. We choose 90 μm for the rest of the experiments.

Two kinds of invariances are studied; shift and rotation. The images stored in the holograms were four faces. The invariance capability was measured by presenting to the network one of the stored images rotated and/or shifted by varying amounts and monitoring the response of the system under various gain conditions. From Fig.9, the minimum gain for this pinhole used is 2.8×10^3 and the maximum gain is 1.2×10^5. We made measurements under low gain ($=3 \times 10^4$) and high gain ($=10^5$) conditions. The results of the shift experiment are shown in Fig.10. Fig.10a shows that as the input image is shifted away from the memory position, the loop response time becomes longer because the correlation signal is shifted away from the pinhole. This makes the initial condition of the loop weaker thus it takes more iterations to reach a stable state. If the input is shifted too much then the correlation peak misses the pinholes completely thus the input is not recognizable. However, the output intensity is shift invariant as long as the loop recognizes the input. Fig.10b shows that the tolerance to shift can be increased by increasing the neuron gain. But in this high gain region the loop has poor discrimination capability and it also incorrectly recognizes a similar face as one of the stored images.

The dynamics and invariance properties under rotation of the input were also measured by using the same pinhole diameter and optical gain. The results are shown in Fig.11. It is seen that by increasing the optical gain from 10^4 to 10^5 the allowable rotation angle for the input is increased from 8 degrees to 16 degrees. Again the dynamics and rotation invariance are consistent with our predictions.

Acknowledgements

This research is supported by the Defense Advanced Research Projects Agency and the Army Research Office. The authors thank Dr. Eung Gi Paek who contributed greatly at the early stages of this work and also David Brady and Claire Gu for many helpful discussions on various aspects of this work.

References

[1] E.G. Paek and D. Psaltis, Optical Engineering, Vol. 26, p.428, 1987.

[2] K. Hsu, D. Brady and D. Psaltis, IEEE NIPS Conf, Denver, 1987.

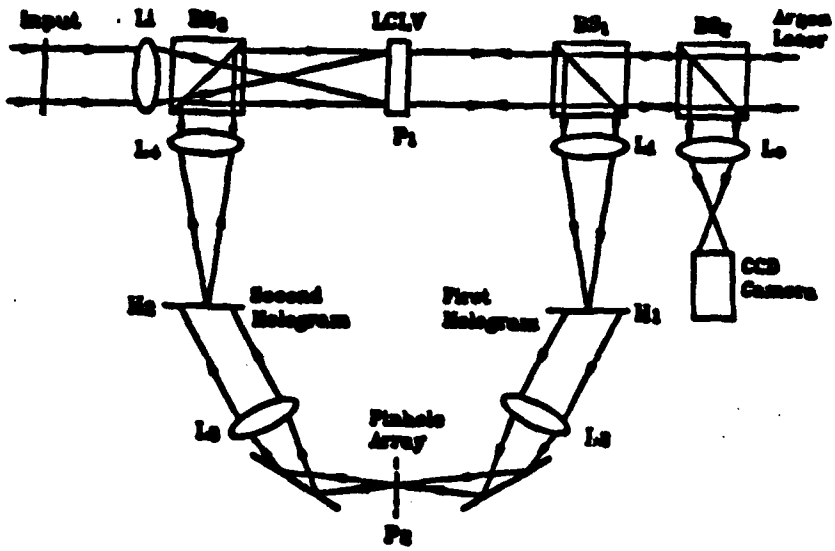


Figure 1. Optical auto-associative loop.

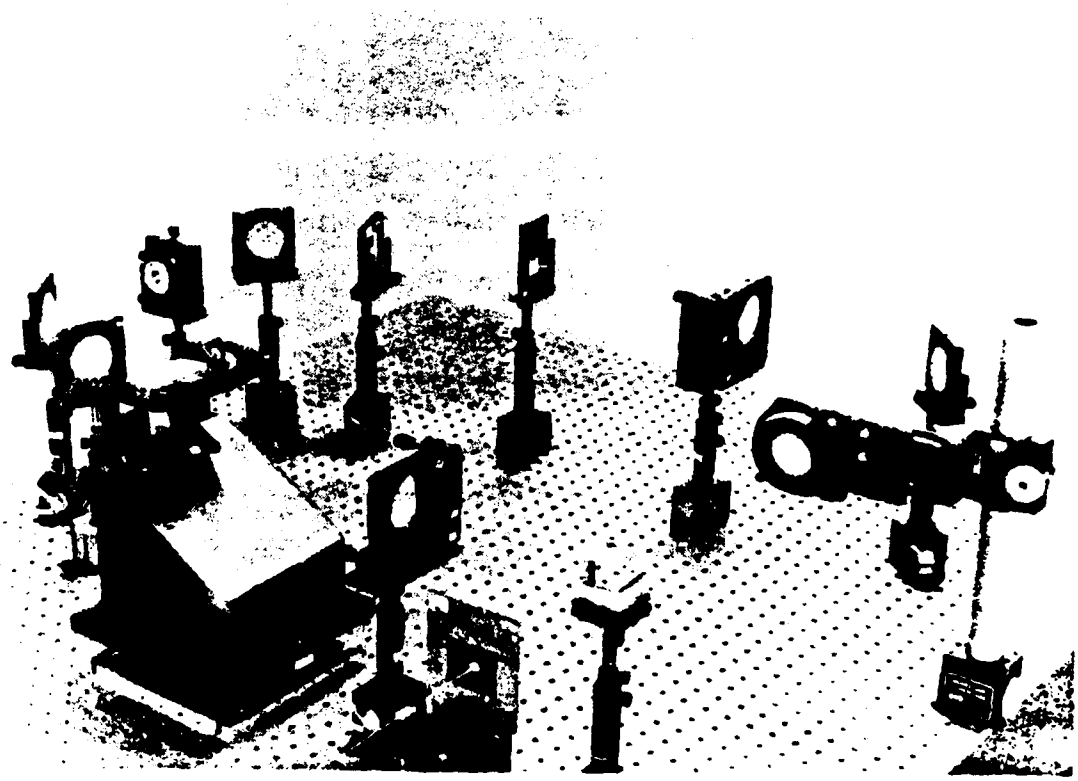


Figure 2. Experimental setup of the optical auto-associative loop.

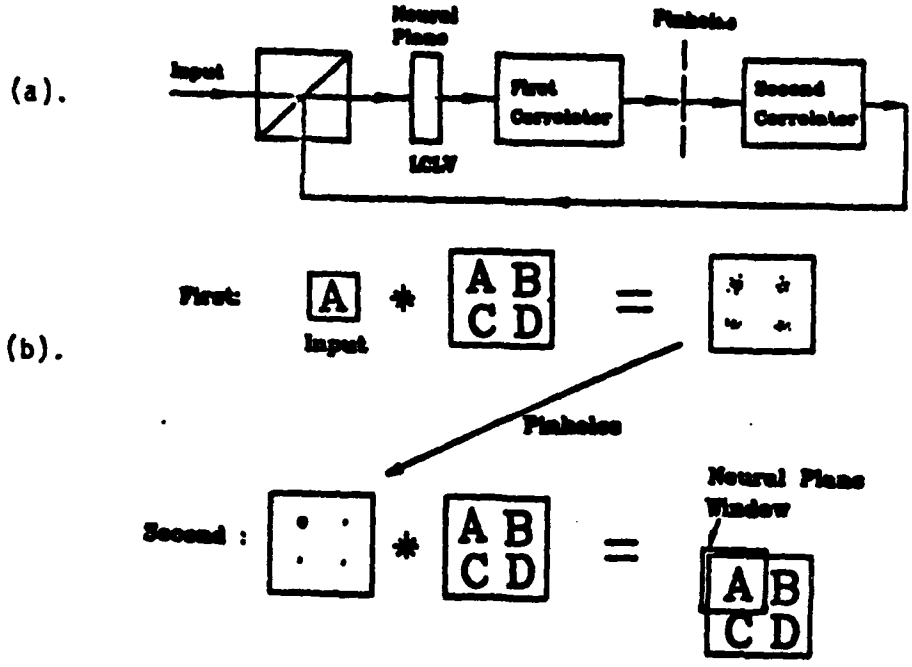


Figure 3. Operation principle of the optical associative loop. (a) Block diagram of the optical loop. (b) Example of recalling one of the stored images from an associative input.

Figure 4. Example of auto-associative recalling. (a) Half-face image input into the loop. (b) Overlap of the input image with the recalled complete image. (c) Stable image in the loop after the input was cut off.

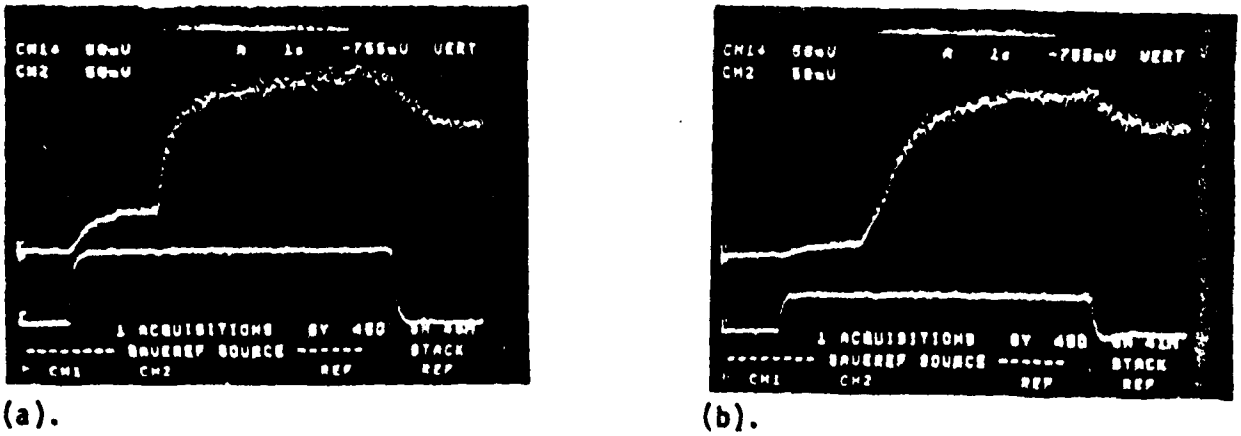


Figure 5. Temporal response of the optical loop. (a) Loop response with a strong initial condition. (b) Loop response with a weak initial condition.

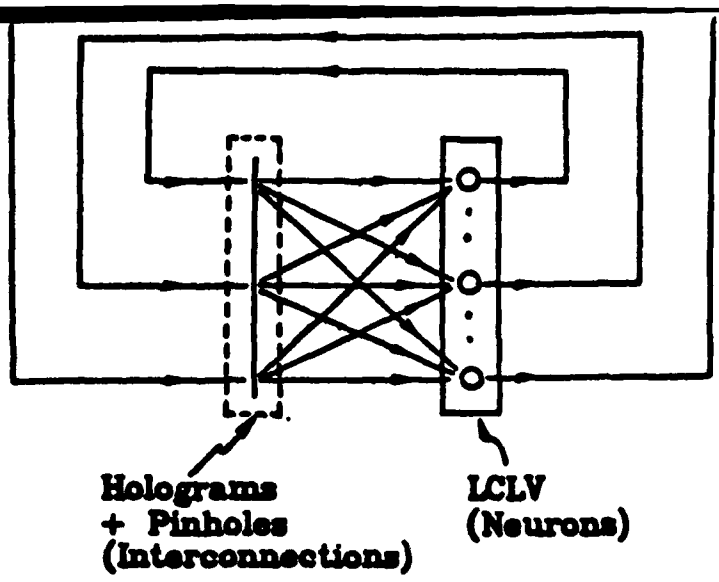


Figure 6. Neural network model of the optical auto-associative loop.

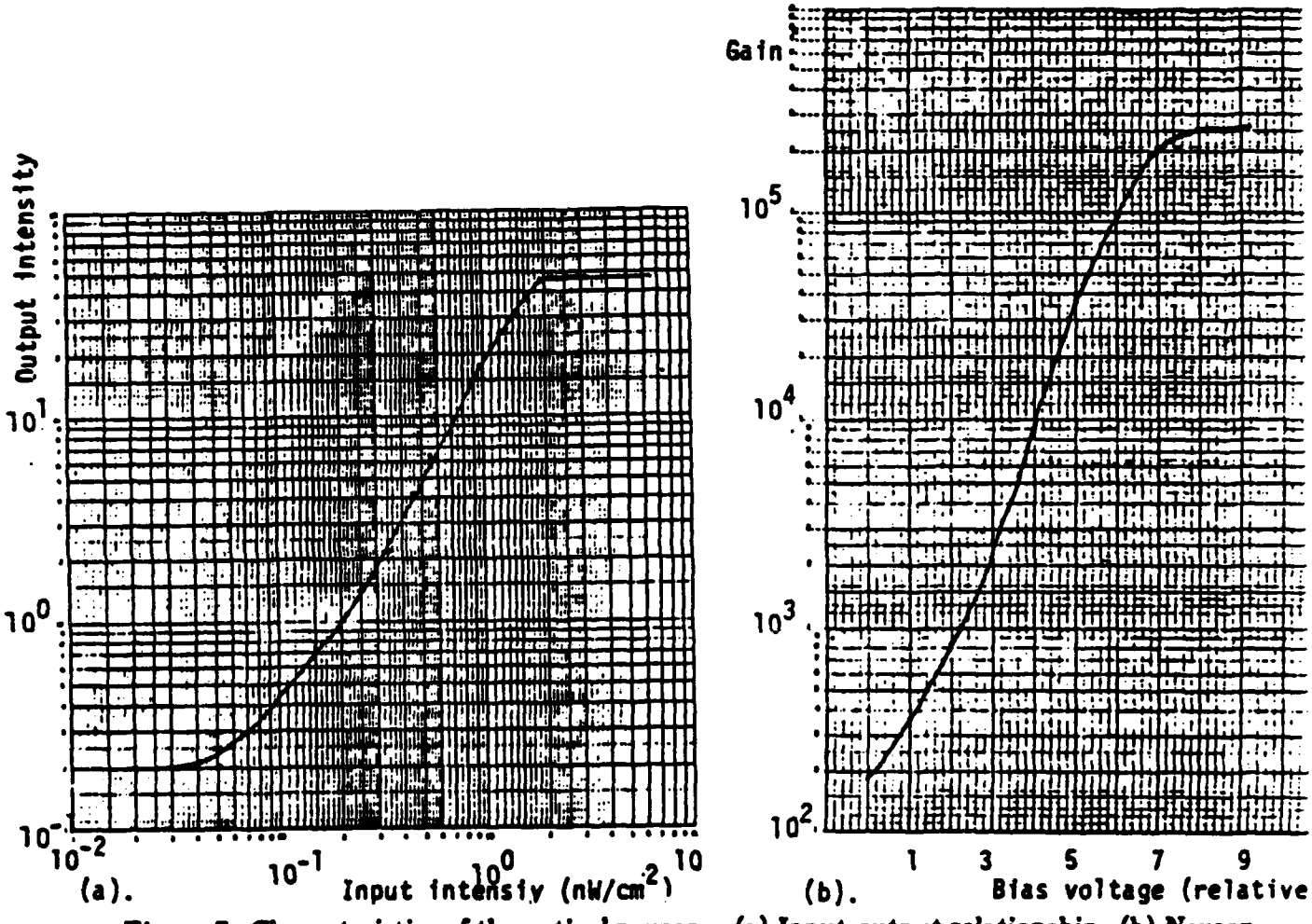


Figure 7. Characteristics of the optical neurons. (a) Input-output relationship. (b) Neuron gain as a function of the bias voltage on the image intensifier. Read-beam = 500 $\mu W/cm^2$.

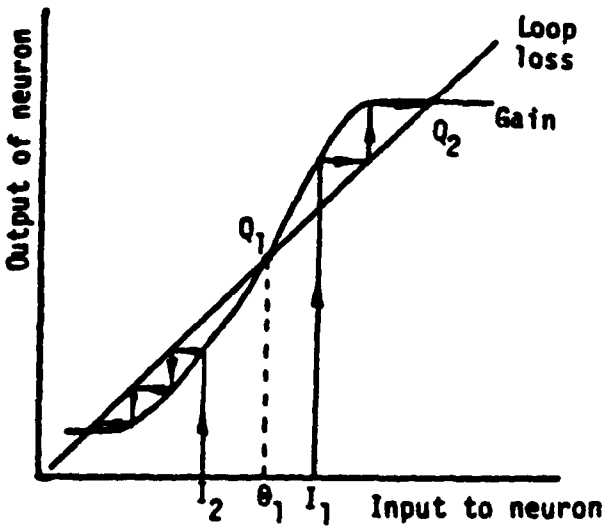


Figure 8. Dynamics of the optical loop. θ_1= threshold value of the loop, I_1, I_2 = initial conditions of the loop.

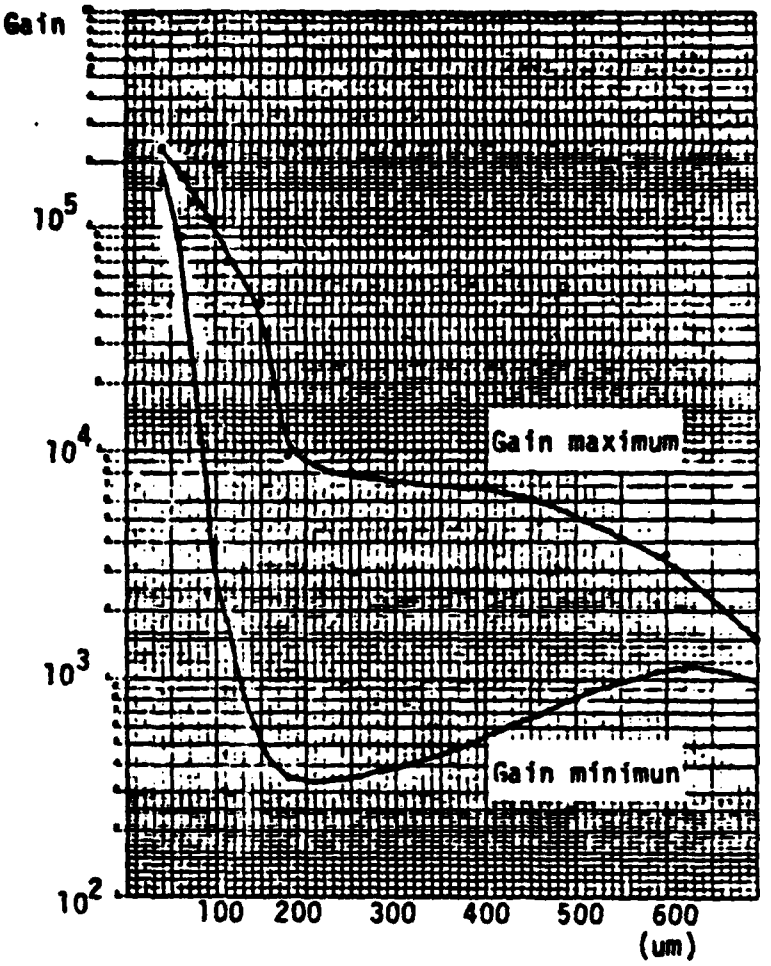


Figure 9. The allowable gain region of the optical loop versus pinhole size.

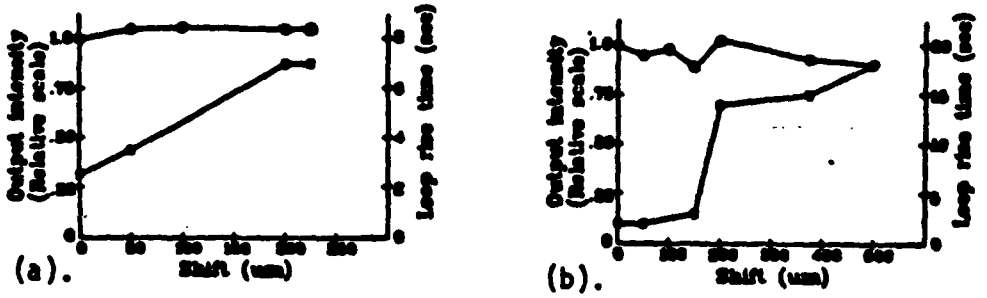


Figure 10. The shift invariance of the optical associative loop. (a) Neuron gain = 3×10^4. (b) Neuron gain = 1×10^5. The upper curve is the loop output intensity, and the lower curve is the loop rise time.

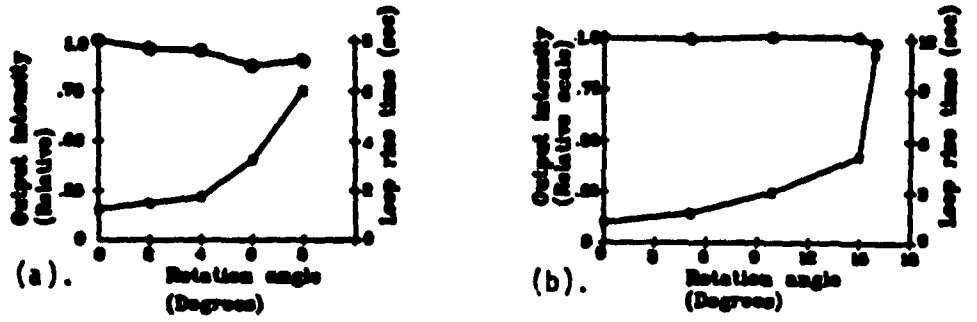


Figure 11. The rotation invariance of the optical asspciativ. op. (a) Neuron gain = 3×10^4. (b) Neuron gain = 1×10^5. The upper curve is the loop output intensity, and the lower curve is the loop rise time.

Reprinted from Applied Optics, Vol. 24, page 3860, November 15, 1985

Bias-free time-integrating optical correlator using a photorefractive crystal

Demetri Psaltis, Jeffrey Yu, and John Hong

An acoustooptic time-integrating correlator is demonstrated using a photorefractive crystal as the time-integrating detector.

I. Introduction

Time integration[1] has proved to be a powerful technique in optical signal processing and has been used in a wide variety of architectures. A major drawback of time-integrating processors is the buildup of bias in addition to the signal. This occurs because the photogenerated charge that is integrated on the detector is proportional to the intensity of the optical signal which makes it necessary to represent bipolar signals on a bias. The effective system dynamic range at the output is given by $DR' = DR\,[SBR/(1 + SBR)]$ where DR is the dynamic range of the output detector and SBR is the signal-to-bias ratio on the detector.[2] In most cases of interest, the SBR is much smaller than unity and thus the added bias significantly reduces the usable dynamic range of the system.

The most frequently used method for separating the signal from the bias involves placing the signal on a spatial carrier and then electronically filtering the output of the integrator. This method of bias removal, however, does not solve the dynamic range problem since the processing is done after the detection of the signal. Also, an additional constraint is placed on the resolution of the detector, since the pixel separation must be less than one-half of the period of the carrier being recorded, which will result in a significant reduction in the available space–bandwidth product at the output.

In this paper a new method for performing time-integrating correlation is described using a photorefractive bismuth silicon oxide (BSO) crystal as the time-integrating element. The correlation is formed on a spatial carrier in the crystal and read out with an auxiliary beam. Since only the signal recorded on a spatial carrier is stored in the photorefractive crystal, the diffracted light that is detected contains the correlation information without the bias. The bias does not reduce the dynamic range of the output detector used for final readout, but rather the diffraction efficiency of the BSO crystal. In addition, the resolution of the BSO crystal is very much higher than that of a CCD detector, allowing the correlation of very high space–bandwidth signals to be formed on a carrier. Finally, since the result of the time-integrating correlator is read out optically, the output can be easily interfaced with other optical systems, thus making new architectural designs possible.

In Sec. II, the theory of optical recording in photorefractive crystal is reviewed and extended to the use of photorefractive crystals as time-integrating elements. The architecture and experimental results are described in Sec. III. Dynamic range, linearity, system limitations, and other performance aspects are discussed in Sec. IV.

II. Photorefractive Crystals as Time-Integrating Optical Detectors

When a photorefractive BSO crystal is illuminated by an intensity grating, electrons are excited from traps into the conduction band. These charges migrate due to diffusion and drift from an externally applied electric field and then recombine in dark regions, creating a spatially varying internal space–charge field. This field modifies the index of refraction in the crystal through the linear electrooptic effect and, as a result, a holographic phase grating is recorded in the crystal. Grating formation in photorefractive media has been extensively studied and modeled.[3,4] We will show here that the photorefractive crystal acts as a time-integrating element.

Let the intensity incident on the crystal be as follows:

The authors are with California Institute of Technology, Department of Electrical Engineering, Pasadena, California 91125.
Received 20 June 1985.
0003-6935/85/223860-06$02.00/0.

$$I(x,t) = \begin{cases} I_0 + \mathrm{Re}[I_1(x,t)\exp(ikx)] \text{ for } t > 0 \\ 0 \text{ otherwise.} \end{cases} \quad (1)$$

Assuming that self-diffraction effects are negligible and that the spatial variations of $I_1(x,t)$ are small compared to the grating frequency k, the intensity of the light that is diffracted when the crystal is illuminated by a readout beam can be shown to be[5]

$$I_{out}(x,t) = \left|\frac{K_1}{\tau}\exp(-t/\tau)\int_0^t \frac{I_1(x,t')}{I_0}\exp(t'/\tau)dt'\right|^2 I_R. \quad (2)$$

I_R is the intensity of the readout beam and K_1 is a complex constant involving the material parameters of the crystal, the grating frequency, and the applied electric field. τ is the complex time constant of the space-charge field and is given by[3] $\tau = K_2/I_0$. K_2 is also a complex constant that depends on the photorefractive material used and the experimental conditions. I_0 is the average light intensity incident on the crystal during exposure.

If $I_1(x,t)$ is expanded into its temporal Fourier components,

$$I_1(x,t') = \int_{-\infty}^{\infty} \tilde{I}_1(x,w)\exp(iwt')dw,$$

then the output intensity for $t \gg \tau$ can be written as follows:

$$I_{out}(x,t) \simeq \left|\frac{K_1}{\tau I_0}\int_{-\infty}^{\infty} \frac{\tilde{I}_1(x,w)}{(1/\tau + iw)}\exp(iwt)dw\right|^2 I_R. \quad (3)$$

The above is recognized to be a low pass filter with cutoff frequency $1/|\tau|$, which is approximately equivalent to the output of a sliding window integrator, with integration time τ. Thus,

$$I_{out}(x,t) \simeq \left|\frac{K_1}{\tau}\int_t^{t+\tau} \frac{I_1(x,t')}{I_0}dt'\right|^2 I_R. \quad (4)$$

Hence, the output intensity can be treated as the square of the normalized integration of the signal I_1. An interesting observation is that, if $I_1(x,t)$ were independent of time, the output intensity depends only on the ratio of the modulated intensity I_1 to the dc intensity I_0.

III. Experimental Demonstration

A schematic diagram of the experimental system is shown in Fig. 1. The input electrical signals are mixed with the center frequency of the acoustooptic devices (AODs) and fed into the AODs. The first AOD is illuminated by a collimated wave and the upshifted diffracted order is imaged onto the second AOD, then reimaged onto the photorefractive BSO crystal. The second AOD is oriented such that the acoustic signal is counterpropagating with respect to the image of the acoustic signal from the first AOD. The undiffracted light transmitted through the first AOD is incident at the Bragg angle of the second AOD. The upshifted diffracted order of the second AOD is also imaged onto the BSO crystal. The undiffracted light is spatially filtered before reaching the BSO crystal. In this arrangement, the AODs are parallel to each other, but the diffracted orders propagate at an angle with respect to each other even though both diffracted beams are temporally upshifted. This causes the signals from the two AODs to interferometrically record the correlation signal on the BSO crystal at a high spatial frequency. Let the inputs to the AODs be $v_1(t) = a(t)\exp(iw_0t)$ and $v_2(t) = b(t)\exp(iw_0t)$, where $w_0/2\pi$ is the center frequency of the AOD. The intensity incident on the photorefractive crystal is

$$\begin{aligned} I(x,t) &= |a(t - x/v)\exp(i\gamma x) + b(t + x/v)\exp(-i\gamma x)|^2 \\ &= |a(t - x/v)|^2 + |b(t + x/v)|^2 \\ &\quad + 2\,\mathrm{Re}[a(t - x/v)b^*(t + x/v)\exp(i2\gamma x)], \end{aligned} \quad (5)$$

where v is the acoustic velocity of the AOD and $\gamma = w_0/v$. We will treat the case where $|a(t)|^2$ and $|b(t)|^2$ can both be approximated as constants, as is the case for FM signals. Then, the intensity pattern can be separated into a dc term I_0 and a signal term $I_1(x,t)$ modulating a spatial carrier $\cos(2\gamma x)$ in the form of Eq. (1). This intensity pattern results in the formation of a hologram on the photorefractive crystal as described in the previous section. The hologram is read out with an auxiliary beam and is imaged onto a charge coupled device (CCD) detector for readout.

If the assumption is valid that $I_1(x,t)$ has spatial frequencies which are small compared with the carrier frequency 2γ, we can use Eq. (4) to obtain an expression for the output intensity detected by the CCD:

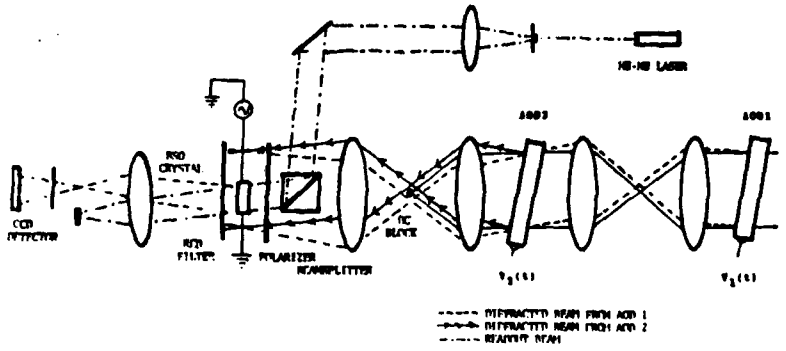


Fig. 1. Optical setup of the photorefractive bias removal correlator.

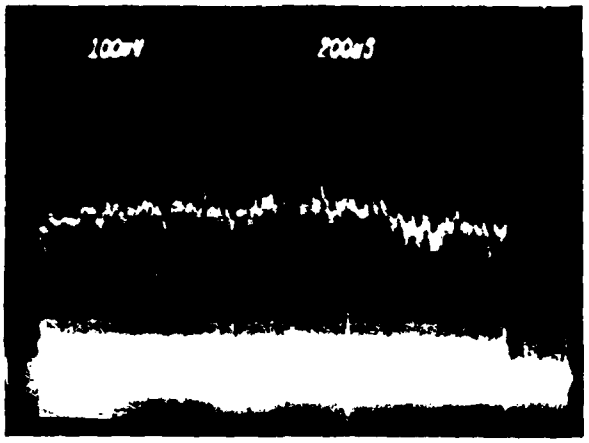


Fig. 2. Output of a standard time-integrating correlator without noise.

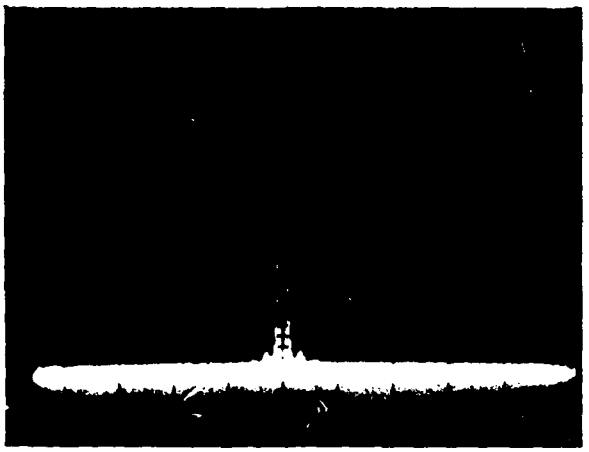


Fig. 3. Output of the bias removal correlator without noise.

$$I_{out}(x) \propto I_R \left|\frac{K_1}{\tau}\right|^2 \left|\int_t^{t+\tau} \frac{a(t - x/v)b^*(t + x/v)}{|a|^2 + |b|^2}\, dt\right|^2,$$

and by defining variable $t_1 = t - x/v$,

$$I_{out}(x) \propto \left|\int_{t+x/v}^{t+\tau-x/v} a(t_1)b^*(t_1 + 2x/v)dt_1\right|^2. \quad (6)$$

Hence the system produces the magnitude square of the correlation between the signals $a(t)$ and $b(t)$ integrated over a finite interval τ.

Flint glass acoustooptic cells driven at a center frequency of 70 MHz were used in the experiment. A symmetric linear chirp signal with bandwidth $\Delta f = 5$ Mhz was fed into each cell to produce the autocorrelation peak. The Bragg angle of the AODs was 0.2°, which corresponded to a grating frequency equal to 35 lines/mm in the BSO crystal.

The BSO crystal used in the experiment was cut in the ⟨110⟩ direction and measured 15 × 15 × 2 mm. An external electric field of 7 kV/cm was applied in the ⟨001⟩ direction of the crystal which was also the direction of the grating vector.

The correlation was recorded on the crystal with an argon laser at a wavelength of 514 nm with average intensity equal to 1 μW/cm^2. The correlation was read out with a He–Ne laser (λ = 633 nm) with 150-μW/cm^2 intensity. Cylindrical lenses (not shown in Fig. 1) were used to expand the output of the AODs thereby illuminating the full aperture of the BSO crystal and also to focus the diffracted light onto a 1-D CCD.

The output signal-to-bias ratio of a conventional time-integrating correlator is reduced when the levels of the two signals are unequal and/or if there is additive noise present in the system. Both conditions were simulated experimentally. Noise was simulated by adding a 70-MHz signal to the input of one of the AODs. The output of a standard time-integrating correlator (i.e., the correlation formed directly on the CCD) for the noise-free case and equal amplitude signals is shown in Fig. 2. This condition provides the maximum signal-to-bias ratio for the system. We can

Fig. 4. Output of the bias removal correlator with a signal-to-noise ratio of 0 dB.

Fig. 5. Output of the bias removal correlator with a signal-to-noise ratio of −10 dB.

see in Fig. 2 that there is still a strong bias term added to the correlation peak. The correlation produced by temporally integrating on the photorefractive crystal is shown in Fig. 3. In this case, all the bias due to temporal integration is removed, and any residual bias is due entirely to dark current from the CCD. The outputs of the bias removal correlator with input signal-to-noise ratios of 0 and −10 dB are shown in Figs. 4 and 5, respectively. Again, bias levels which appear in the figures were entirely due to the integration of dark current in the output detector. In practice, the detector dark current can be minimized by increasing the intensity of the readout beam, thereby decreasing the required integration time of the output CCD detector and/or cooling the detector.

III. Performance

The experimental results described in the previous section show a dramatic qualitative improvement in the correlation that is obtained when the photorefractive crystal is used instead of the CCD. In this section we examine certain characteristics of this method which are useful for quantitatively evaluating its performance. Specifically, we examine the linearity, integration time, dynamic range, and sensitivity of the correlator.

A. Linearity

In a conventional time-integrating correlator (coherent or incoherent), the output correlation is basically proportional to the signals applied to the AODs. Nonlinearities occur only when we exceed the linear dynamic range of the devices used, i.e., if the diffraction efficiency of the AOD exceeds several percent or the integrating detector is driven to saturation. In the photorefractive time-integrating processor, the output intensity is a nonlinear, monotonically increasing function of the input voltage. The nonlinearity arises because of the square-law detection at the final readout stage and the recording mechanism in the photorefractive crystal. The nonlinear relationship is now studied analytically and experimental verification of the theoretical results is presented.

Let $v_1(t) = s(t)$ be a fixed reference signal and $v_2(t) = as(t)$ be an input signal of varying amplitude ($0 < a < 1$). Since the correlation term contains spatial frequencies which are much lower than the grating frequency, near the correlation peak ($x = 0$) the intensity incident on the photorefractive crystal is

$$I(x,t) = (1 + a^2 + 2a\cos kx)|s(t)|^2.$$

Using Eq. (1), the output intensity at the CCD is proportional to

$$I_{out} \propto \left|\frac{2a}{1+a^2}\tau\right|^2.$$

The modulation depth of the intensity incident on the BSO crystal is

$$m = \frac{2a}{1+a^2},$$

and hence

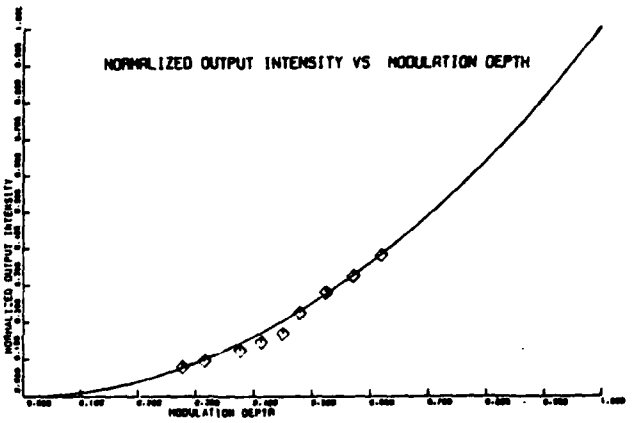


Fig. 6. Normalized output intensity vs modulation depth.

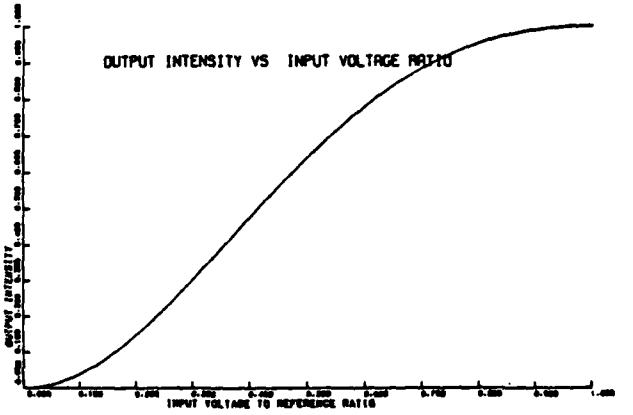


Fig. 7. Theoretical plot of output intensity vs input voltage ratio.

$$I_{out} \propto m^2 = 4a^2/(1+a^2). \quad (7)$$

Figure 6 is a graph of the output intensity at the correlation peak vs the modulation depth incident on the crystal. The experimental result is in excellent agreement with the square-law relationship predicted by Eq. (7).

A plot of the output intensity as a function of the amplitude of the input signal a is shown in Fig. 7. The nonlinear relationship between the input and output signals is generally a disadvantage since the scaling of signals of varying amplitudes will be nonlinear. This, however, will not cause a problem if the correlator is used only as a signal detection device, since correlation peaks will still be discernible and only the threshold level need be adjusted accordingly to maximize the probability of detection.

B. Integration Time

In a conventional time-integrating correlator, the integration time is limited by the dark current buildup on the output detector, typically up to several hundred milliseconds. When the photorefractive crystal is

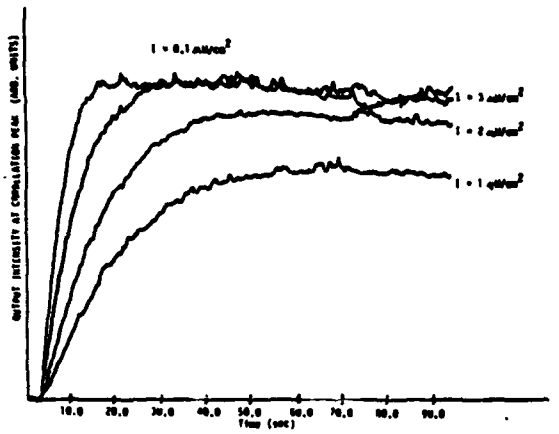


Fig. 8. Output intensity at correlation peak vs time as a function of different average incident intensities.

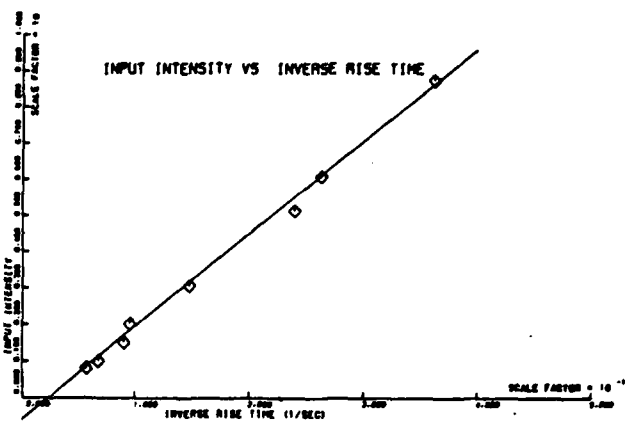


Fig. 9. Inverse of the rise time vs average incident intensity.

used, the integration time is determined by the rise time of the internal space–charge field which can easily be made much longer. The correlation can be read out at any rate that is convenient by an auxiliary detector array.

The integration time is approximately equal to $|\tau|$, where

$$|\tau| = \left|\frac{K_2}{I_0}\right| . \tag{8}$$

Hence, the integration time of the bias removal correlator can be controlled by varying the writing intensity. This control is important since the integration time can be matched to the length of the reference signal thereby increasing the probability of detection of a weak signal.

The time response of the correlation peak for different values of average incident intensity is shown in Fig. 8. Figure 9 is a plot of intensity vs the inverse of the experimentally observed rise time. There is excellent agreement between the experiment and Eq. (8).

The integration time, however, has a finite range over which it can be adjusted. The maximum integration time is limited by the thermal effects in the crystal. If the rate at which carriers are generated thermally becomes comparable with the rate at which they are photogenerated, the modulation depth of trap density will be reduced. As a result, the diffraction efficiency of the grating will decrease. In practice, the minimum integration time is limited by the maximum light intensity that is available for recording. The integration time can be reduced to 30 msec if the incident intensity is made equal to 18 mW/cm^2. This power level, however, is simply not practical for most applications.

C. Dynamic Range and Sensitivity

Since the output of the bias removal correlator is presented without bias, the output dynamic range of the system is essentially equal to the dynamic range of the readout detector array. To characterize the performance of the system we need to determine how the input signal levels are mapped to this output dynamic range. Let the dynamic range of the photorefractive crystal be defined as $DR_{BSO} = m_{max}/m_{min}$, where m_{max} is the maximum modulation depth ($m_{max} = 1$), and m_{min} is the minimum modulation depth for which a diffracted signal is detectable above the output scatter and noise level of the system.

Given two input signals $v_1(t) = as(t)$ and $v_2(t) = s(t)$, the modulation depth of the light incident on the crystal is $m = 2a/(a^2 + 1)$. Thus, the minimum detectable input signal is given by $a_{min} = m_{min}/2 = 1/DR_{BSO}$. The useful range over which a can vary is limited by DR_{BSO}. From a_{min}, one can define an input dynamic range given by $DR_{input} = 1/a^2_{min} = 4/m^2_{min}$. The most important parameter in determining the system dynamic range is the minimum detectable modulation depth m_{min}. Experimentally, we measured the dynamic range to be equal to 23 dB. This corresponds to a minimum modulation depth of 0.142. We expect that through careful design this can be substantially improved. However, all the mechanisms that determine m_{min} are not fully understood. It is believed that besides detector noise and scattering from the crystal, the modulation depth is limited by thermal effects in the material and shot noise arising from the internal currents.

Another important aspect of the correlator system is its sensitivity or the minimum signal-to-noise ratio that is detectable. This parameter is also determined by the minimum detectable modulation depth, m_{min}. Given a reference signal $v_1(t) = as(t)$ and an input signal contaminated by additive noise, $v_2(t) = bs(t) + n(t)$, the modulation depth of the intensity incident on the crystal is

$$m = \frac{2ab|s(t)|^2}{(a^2 + b^2)|s(t)|^2 + \sigma_n^2} .$$

The reference level which maximizes m is given by $a = (b^2 + \sigma_n^2/|s(t)|^2)^{1/2}$, corresponding to a modulation depth of

$$m = \frac{b}{(b^2 + \sigma_n^2/|s(t)|^2)^{1/2}}.$$

In practice, optimizing the reference level can easily be achieved by setting the power of the reference equal to the total average power of the input signal.

Normalizing the signal and noise terms such that $|s(t)|^2 = \sigma_n^2 = 1$ we obtain

$$m = \frac{b}{(b^2 + 1)^{1/2}}.$$

Thus, the minimum input SNR that produces a detectable correlation peak at the output is $(S/N)_{min} = (b^2/\sigma_n^2)_{min} \cong m_{min^2}$.

From the experimentally measured value of m_{min}, the correlator should have had a sensitivity of −17 dB. However, experimental results showed a sensitivity of −14 dB.

IV. Conclusion

The photorefractive time-integrating processor that has been described has several advantageous features: bias removal, increase in the output space–bandwidth product, and the ability to directly interface the result of the time-integrating processor with other optical systems. Bias-free correlation is desirable because it allows us to increase the dynamic range and hence the sensitivity of time-integrating processors. In the implementation described in this paper, however, the square-law detection at the output reduces the available overall dynamic range. A definite improvement in dynamic range can be obtained if the correlation that is formed in the photorefractive crystal is interferometrically detected on the output detector. Another limitation of the system described here is the long integration time (several seconds). In some applications this long integration time is desirable and could result in extremely good sensitivity (detection of signals with very low SNR). However it is certainly desirable to be able to decrease the integration time to several milliseconds. This could be accomplished by increasing the optical power of the writing beams, but this is in general an impractical solution. Another limitation of this technique is the relatively low diffraction efficiency that is obtained with BSO crystals (2–3%), which reduces the overall light efficiency. Materials with higher electrooptic coefficients, such as barium titanate, can provide better efficiency; however, the time constant obtained with this particular material is much longer than that of BSO. New photorefractive materials currently being developed showing promise of a large improvement in optical sensitivity as well as higher electrooptic coefficients may provide a substantial improvement in performance and, specifically, reduce the total optical power that is required.

This work is supported by the Air Force Office of Scientific Research and the Army Research Office.

References

1. R. A. Sprague and C. L. Koliopoulos, "Time Integrating Acoustooptic Correlator," Appl. Opt. 15, 89 (1976).
2. D. Psaltis, "Incoherent Electrooptic Image Correlator," Opt. Eng. 23, 12 (1984).
3. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, and M. S. Soskin, "Holographic Storage in Crystals. I: Steady State," Ferroelectrics 22, 949 (1979).
4. J. Feinberg, D. Heiman, A. R Tanguay, and R. W. Hellwarth, "Photo Refractive Effects and Light-Induced Charge Migration in Barium Titanate," J. Appl. Phys. 51, 1297 (1980).
5. M. Cronin-Golomb, "Large Nonlinearities in Four-Wave Mixing in Photorefractive Crystals and Applications in Passive Optical Phase Conjugation," Ph.D. Thesis, California Institute of Technology (Mar. 1983).

Submitted to Journal of Applied Physics

VOLUME HOLOGRAPHIC INTERCONNECTIONS WITH MAXIMAL CAPACITY AND MINIMAL CROSS TALK

Hyuk Lee

Polytechnic University

Department of Electrical Engineering

Brooklyn, New York 11201

Xiang-guang Gu and Demetri Psaltis

California Institute of Technology

Department of Electrical Engineering

Pasadena, California 91125

ABSTRACT

Optical interconnections utilizing volume holography is described. Intrinsic cross-talk effects that limit the number of independent interconnections are identified and analyzed by applying coupled-wave analysis. Sampling grids for removing the first-order cross talk are presented resulting in a system limited by second and third order cross talk only.

INTRODUCTION

Optical interconnecting elements that exploit free propagating light waves can potentially act as a powerful alternative to electrical wiring because free propagating photons lack the interactive nature of electrons [1]. Optical interconnections can be particularly useful for the optical implementation of neural computers [2] in which each processing element is interconnected to many others (typically several thousand). As an example, a network that is capable of processing images may consist of several million processing elements (or "neurons") and therefore there is a very large number of interconnections to be specified in such a system. If the system that is utilized to simulate all these connections is planar (e.g. electronic or an optical system that utilizes a planar medium to specify the connectivity pattern), then the area of the device grows in proportion to the total number of connections. As an example, let us assume that the area required to record the strength of each interconnection is $10\mu m^2$, then the total area required to simulate a network that is comprised of 10^9 connections is 10cm $\times$ 10cm. This makes the fabrication of such a device very difficult and in the case of the optical implementation, the size of the optical system becomes exceedingly large. To overcome this shortcoming, we have previously proposed [3] a holographic optical interconnection method for utilizing a three-dimensional storage medium which provides a much higher storage density. In this paper, we derive the interconnection pattern having minimum cross talk and the signal-to-noise ratio for this interconnecting configuration.

HOLOGRAPHIC INTERCONNECTIONS

To identify the fundamental cross-talk effects that limit the available number of independent interconnections, a global volume holographic interconnection between N input and N output pixels is considered. The arrangement we will be using is shown in Fig. 1. The input and output pixels are arranged in planes. A lens collimates light from each input point and therefore the light incident on the crystal in Fig. 1 due to a single point at the input is a plane wave whose propagation direction is determined by the position of the pixel. Similarly, an output lens focuses each diffracted plane wave to a pixel on the output plane. The interconnection between each pair of input-output points is performed by a separate grating, with the strength of each grating determining the weight of the connection. Each grating can be recorded with a separate exposure which would require a total of N^2 exposures. We can reduce the number of required exposures by forming N multiple holographic exposures [4] as follows. One input point is turned on during each exposure and the desired connectivity pattern between the selected input point and all the output points is recorded at the training plane (see Fig. 1). An exposure of the interference pattern between the two waves is recorded and the process is repeated for each of the N input points. If we neglect diffraction effects at the crystal boundaries, then the interconnection pattern consists of perfect sinusoidal gratings, which include: (1) $N(N-1)/2$ gratings that are recorded by the interference between pixels that are simultaneously on at the training plane during the recording, and (2) N^2 gratings connecting input and output pixels. For convenience, the former set of $N(N-1)/2$ gratings are referred to as intra-layer gratings and the latter set of N^2 gratings are described as inter-layer gratings.

An independent interconnection is defined in such a way that the intensity I_p of the diffracted light wave at the output pixel p is given by

$$I_p = \sum_{i=1}^{N} \eta_{pi} I_i, \tag{1}$$

where i is the index that represents input pixels, I_i is the intensity at the input pixel i, and η_{pi} is the diffraction efficiency of the grating generated by the interference between the input pixel i and the pixel p' at the training plane that corresponds to the output pixel p. We have assumed in Eq. (1) that the read-out light is spatially incoherent. This means that the light intensity reaching each output point is a linear combination of the light *intensities* of the input pixels and therefore we have used light intensity as the variable that describes the system. If the hologram is read-out with spatially coherent light, then the field is the appropriate variable to use. The field is a complex quantity (has both amplitude and phase) and therefore the coherent case is generally more difficult to analyze and also implement.

The cross-talk effect in volume holographic interconnections is defined as the difference between the actual light intensity $\tilde{I}_p$ obtained at the pixel p and the desirable intensity of Eq. (1). If we consider only first-order cross talk (i.e. neglecting the contribution of multiple diffraction) we can write $\tilde{I}_p$ as follows.

$$\tilde{I}_p = \sum_i \eta_{pi} I_i + \sum_{j \neq i} \sum_{l \neq p} \sum_i \eta_{ipjl} I_i, \tag{2}$$

where η_{ipjl} is the diffraction efficiency with which light is diffracted from pixel i to pixel p due to a grating that was recorded by pixel j at the input and l at the output. Coupled-wave analysis [5] is utilized below to evaluate those cross-talk effects. For this purpose small

diffraction efficiency for an individual grating is assumed. Due to the assumption of small diffraction efficiency, the primary cross-talk effect can be evaluated by means of a first-order coupled-wave analysis. In such an analysis, an input light wave at pixel i interacts with every grating in the volume hologram independently and without an intermediate rediffraction. Diffracted light waves from the intra-layer gratings do not contribute to cross-talk effects because of large phase mismatch, and therefore we only need to consider the first-order cross talk that results from the inter-layer gratings.

Let us consider an output pixel p. The light intensity received at p including the first-order cross-talk effect is given by Eq. (2). The cross-talk diffraction efficiency η_{ipjl} calculated from coupled mode analysis is approximately [5]:

$$\eta_{ipjl} = \begin{cases} \eta_{jl} sinc^2(\Delta k_{ipjl} L/2\pi) & \text{if } (2\pi/\lambda)\mathbf{n}_i + \mathbf{K}_{jl} = \mathbf{n}_p \| (2\pi/\lambda)\mathbf{n}_i + \mathbf{K}_{jl} \| \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $\mathbf{n}_i$ and $\mathbf{n}_p$ denote unit vectors in the direction of propagation from the input pixel i and towards the output pixel p, respectively. L is the thickness of the crystal and Δk_{ipjl} denotes the phase mismatch for the interaction between the grating $\mathbf{K}_{jl}$ that has been recorded for interconnecting point j to point l and the optical wave emanating from the input pixel i and it is given by

$$\Delta k_{ipjl} = \| (2\pi/\lambda)(\mathbf{n}_i - \mathbf{n}_p) + \mathbf{K}_{jl} \|. \tag{4}$$

λ is the optical wavelength in the crystal.

FIRST ORDER CROSS TALK

The first-order cross talk can be eliminated if one can arrange input and output pixels

so that all the N^2 gratings in Eq. (2), except for the grating $\mathbf{K}_{ip}$ yielding the signal, belong to one of the two types of gratings defined below. The first type is characterized by the condition that the phase mismatch given by Eq. (4) is larger than $2\pi/L$, in which case the diffracted light intensity is very small; for these gratings, the first-order cross-talk effect can be neglected. The second type consists of gratings for which the diffracted light waves do not propagate to any one of the output pixels used for the interconnection in which case from Eq. (3) we have that $\eta_{ipjl} = 0$ and hence such gratings do not contribute any light intensity at the pixel p through first-order cross talk. To derive an arrangement that will ensure that all the recorded gratings satisfy one of the two conditions stated above, we note that the phase mismatch described by Eq. (4) is determined by the geometry of the input and output pixels. The wave vector diagram is drawn in Fig. 2, where $\mathbf{k}_i$ and $\mathbf{k}_j$ refer to the input wave vectors and $\mathbf{k}_p$ and $\mathbf{k}_l$ are the output wave vectors. The condition

$$\mathbf{n}_p = \frac{\mathbf{k}_i + \mathbf{K}_{jl}}{\|\mathbf{k}_i + \mathbf{K}_{jl}\|} \tag{5}$$

states that the unit vector $\mathbf{n}_p$ is in the direction of the vectorial sum of the input vector $\mathbf{k}_i$ and the grating vector. This indicates that the grating $\mathbf{K}_{jl}$ is a grating of the first type, being capable of diffracting light from i to p unless it is phase mismatched. Therefore, once a pair (i, p) is selected it is imperative that all the remaining points (j, l) are selected such that if Eq. (5) is satisfied then Δk_{ipjl} is bigger than $2\pi/L$. The degeneracy condition that must be avoided is

$$\Delta k_{ipjl} = \|\mathbf{k}_i + \mathbf{K}_{jl} - \mathbf{k}_p\| < 2\pi/L \tag{6}$$

This condition specifies two strips on the k-space sphere as shown in Fig. 2. The two strips are parallel circles on the wavenormal sphere. The planes in which the strips lie are

perpendicular to the grating vector $\mathbf{K}_{jl}$. If we select an input-output pair (i,p), then if an additional input point j is outside the bottom strip in Fig. 2 it will not produce cross talk to point p; if j is within the bottom strip then cross talk will be eliminated if an output point is not placed at the same location as l along the top strip. If these two criteria can be met for all input and output pixels, then first-order cross talk is completely eliminated. The required width of the strip in Fig.2 is determined by several factors including diffraction due to transverse aperture of the hologram, an effect we have not considered in this paper. The principal factor determining the width is the angular sensitivity of diffraction from a thick grating which is determined by the thickness of the crystal. The width of the strip that is required to satisfy Eq. (6) can be approximated for the purposes of this simplified exposition by $2\pi/Lsin\theta$, where θ is the angle between $\mathbf{k}_i$ and $\mathbf{k}_p$. This estimate is found by determining the angular deviation of the incident and diffracted from the ideal Bragg condition, that will make $\eta_{iplj} = 0$ (see Eq.(3)).

In the above discussion we have specified the conditions that must be met so that each grating implements an independent interconnection in the crystal. The remaining task is to specify the arrangement of input and output pixels in the geometry of Fig. 1 so that the stated conditions are satisfied. We have developed an entire family of sampling patterns that accomplish this goal [3]. Shown in Fig. 3 is one such sampling pattern for the input and output planes. To see why this is the case consider first the gratings connecting two input points along the same row to two points at the same row at the output. These gratings can never be parallel to each other (i.e. fall within the same strip) because the horizontal (x direction in Fig.1) *difference* in position between the input and the output

locations is guaranteed to be different. If we consider two adjacent points in the same column at the input being connected to two adjacent points in the same column at the output, then we find that the two gratings connecting them are tilted with respect to each other in the $y-z$ plane (see the geometry of Fig. 1). In general, gratings connecting points that are neither at the same row or column have gratings that are tilted with respect to each other in all three directions. The patterns in this example are drawn on a $9 \times 9 = 81$ rectangular grid and only $9^{3/2} = 27$ points are utilized as input and output points in the input and output planes, resulting in a total number of connections $9^3 = 729$. In general, if the number of points available on a 2-D rectangular grid is S^2, then the number of pixels that are used for placement of neurons must be $N \leq S^{3/2}$ in order to ensure that the first-order cross talk can be eliminated. Equivalently, if we wish to have N units in the input or output plane, then the number of resolvable points available must be $S^2 = N^{4/3}$.

HIGHER ORDER CROSS TALK

Second-order cross-talk effects result from light waves that are first diffracted by a grating from an input wave at pixel i, and then rediffracted by a second grating and is directed to the output pixel p. Therefore, two gratings are needed. All the second-order light waves resulting from diffraction by two intra-layer gratings or two inter-layer gratings are negligible because in the geometry of Fig. 1 they are phase mismatched and thus they do not contribute to second-order cross-talk effects. Therefore, the principal source of second-order cross talk is diffraction from the inter-layer gratings followed by rediffraction from the intra-layer gratings. Consider again an output pixel p receiving light from an input

pixel i not directly through diffraction by the grating $\mathbf{K}_{ip}$ but through the intermediate step of diffraction first by an inter-layer grating $\mathbf{K}_{ij}$ followed by rediffraction by the grating $\mathbf{K}_{jp}$. This is depicted in a k-space diagram in Fig. 4a. Each input wave diffracts light to all $N^{4/3}$ output pixels through inter-layer gratings and at least N pixels are exactly Bragg matched to the p-th pixel through rediffraction of the intra-layer gratings. Assuming that the overall diffraction efficiency is small and therefore neglect the depletion of the incident beam, we can easily calculate the second-order signal to noise ratio (SNR_2), defined as the ratio of the intensity received at each output pixel due to the direct, first-order diffraction, divided by the total intensity received due to double diffraction:

$$SNR_2 = \frac{\sum_i^N \eta_{pi} I_p}{\frac{1}{4}\sum_i^N \sum_{j\neq p}^N \eta_{ij}\eta_{pj} I_p} \approx \frac{4}{N\eta_1}. \tag{7}$$

In the above equation η_1 is the average diffraction efficiency for an intra-layer grating. From Eq. (7) we see that it is desirable to minimize the strength of the intra-layer gratings to eliminate the second-order cross talk. This can be accomplished by selecting a holographic recording medium in which low spatial frequencies are recorded weakly. This is for instance typical of gratings recorded in photorefractive crystals in the absence of an applied electric field, in which case the recording is done principally by diffusion of the carriers. In this case, gratings whose period is considerably longer than the diffusion length are not recorded effectively. As an example, if $KNbO_3$: Fe 300 ppm is utilized [6], the diffraction efficiency for $.3\mu m$ fringe spacing is more than three orders of magnitude larger than the diffraction efficiency for a fringe spacing of $2.6\mu m$. Hence if the arrangements of input and output pixels are chosen such that the spatial frequency of the inter-layer gratings is much higher than that of the intra-layer gratings, then the effects of intra-layer gratings can potentially

be made negligible compared to third-order cross-talk effects, which we consider next.

Third-order cross talk arises when light originating from the i-th pixel is diffracted by three separate inter-layer gratings and is ultimately directed at the output pixel p. In order to calculate the total amount of third-order cross talk we need to determine the total number of three Bragg matched inter-layer gratings whose vectorial sum is equal to $\mathbf{K}_{ip}$. An example of this condition is depicted in Fig. 4b. The input beam in the direction of the i-th pixel is Bragg matched to $N^{4/3}$ gratings [7] and similarly, a beam diffracted towards the l-th output pixel is Bragg matched (and therefore rediffracted by) $N^{4/3}$ gratings. The ratio of the intensities due to first and third-order diffraction is

$$SNR_3 = \frac{\sum_i^N \eta_{pi} I_i}{\frac{1}{36}\sum_i^N \sum_{l\neq p}^{N^{4/3}} \sum_{j\neq i}^{N^{4/3}} \eta_{li}\eta_{lj}\eta_{pj} I_i} \approx \frac{36}{N^{8/3}\eta_2^2}, \tag{8}$$

where η_2 is the average diffraction efficiency of an inter-layer grating. The conclusion that we might draw from Eq.(8) is that as the network becomes larger (i.e. N increases) the signal-to-noise-ratio deteriorates and therefore third order cross talk imposes a limit on N. In fact, $\eta_2 = \eta_0/N^2$ [4] where $\eta_0 \approx 1$ is the diffraction efficiency obtained when only a single grating is recorded in the crystal. Substitution into Eq.(8) reveals that SNR_3 is proportional to $N^{4/3}$ which implies that for large networks third order crosstalk is not expected to be a serious concern.

CONCLUSION

We have used coupled mode analysis to derive a simple, approximate result for the conditions that must be met in order for each grating that is recorded in a volume hologram to implement an independent interconnection between two points in space. Since the

number of gratings that can be stored in a volume medium is in the order of V/λ^3 [8] where V is the volume of the crystal and λ is the wavelength, the result reported here can make possible the design of optical networks with extremely high storage density. The effects of second and third-order diffraction were calculated and it is shown that these effects can impose a limit on the number of units that can be interconnected with the same crystal, since the signal to noise ratio decreases monotonically as N increases. There are of course several other factors, beyond the basic geometric constraints treated in this paper, which need to be taken into consideration in order to gain a complete understanding of the capabilities of volume holograms for implementing global interconnections. Most significantly, the effects of the recording mechanism and the limitations it imposes on the number of interconnections that a single hologram can implement [4] must be addressed and combined with the results reported here. This will be the subject of a future publication.

ACKNOWELEDGEMENTS

This work was supported at Caltech by the Air Force Office of Scientific Research, the Army Research Office and DARPA, and at Polytechnic by the National Science Foundation EET-8810288 and the Center for Advanced Technology in Telecommunications.

REFERENCES

[1] J. W. Goodman, F. J. Leonberger, S. Y. Kung and R. A. Athale, Proc. IEEE, 72, pp. 850-866, 1984.

[2] D. Psaltis and N. Farhat, Opt. Lett. 10, 98, 1985.

[3] D. Psaltis, X. Gu, J. Yu and H. Lee, Tech. Digest, Topical Meeting Opt. Comp., Lake Tahoe, pp. 129-132, 1987.

[4] D. Psaltis, D. Brady and K. Wagner, Appl. Opt. 27, 1752, 1988.

[5] H. Kogelnik, Bell Syst. Tech. J. 48, 2909, 1969.

[6] P. Gunter, Phys. Reports, 93, pp. 199-299, 1982.

[7] D. Psaltis, X. Gu, H. Lee and J. Yu, submitted to Applied Optics

[8] P. Van Heerden, Appl. Opt. 2, pp. 387-392, 1963.

FIGURE CAPTIONS

Fig. 1 Illustration of the proposed interconnection scheme between an input point i and an output point p. One grating is stored by interfering the two beams coming from point sources i and p' after passing through a Fourier transforming lens. Point p' is the inverted image of point p. After storing the grating, light coming from point i is diffracted by the grating and focused on point p. Therefore, the stored grating interconnects points i and p.

Fig. 2 k-space diagram illustrating the degeneracy of the gratings that connect points (i, p) and (j, l).

Fig. 3 Sampling patterns on 9×9 rectangular grids.

Fig. 4 Wave-vector matching diagram illustrating the mechanism through which a) second and b) third order cross talk is introduced at each output pixel p.

Fig. 1

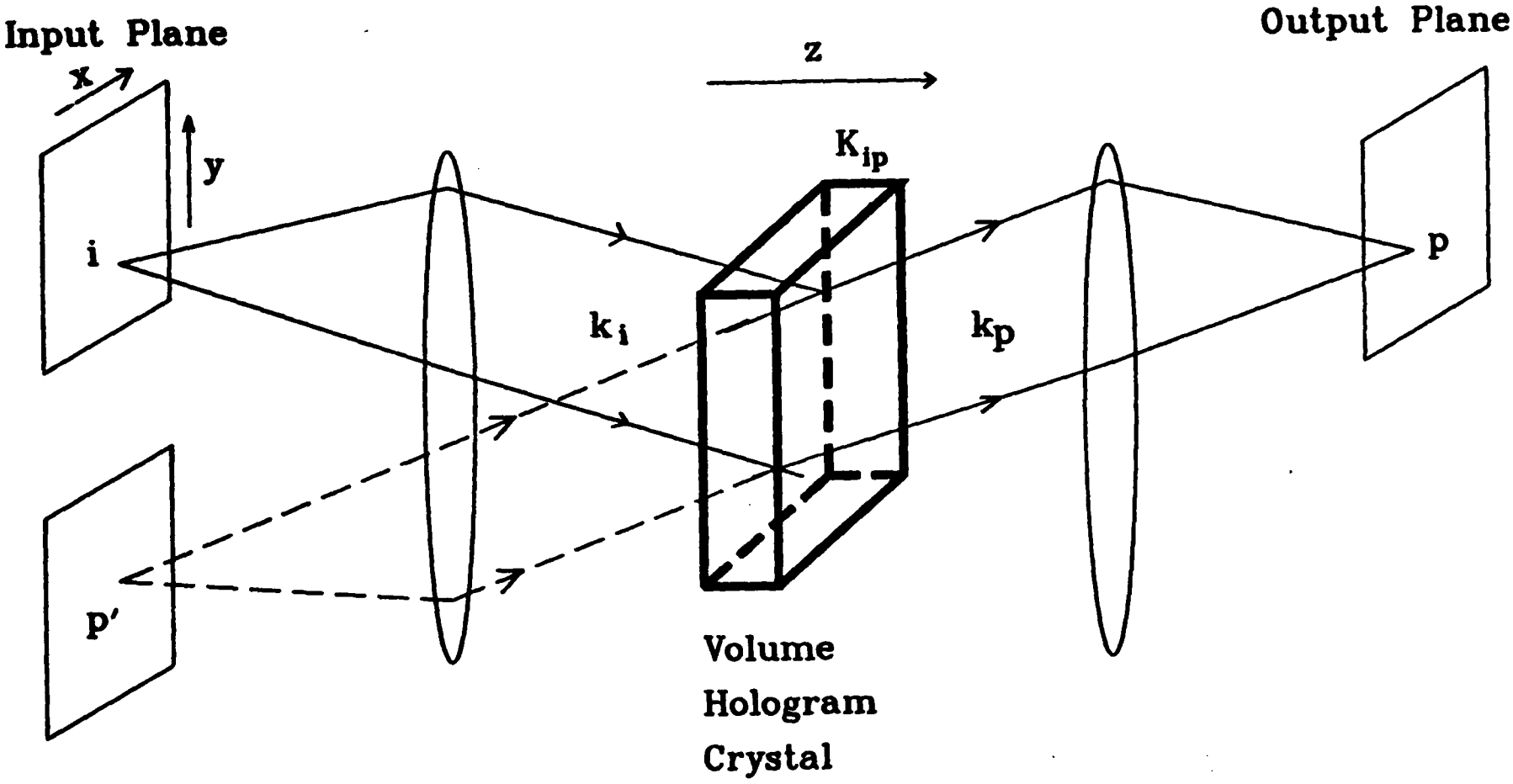

Input Plane
x
y
i
z
K_{ip}
k_i
k_p
Output Plane
p
p'
Volume
Hologram
Crystal
Training Plane

Fig. 2

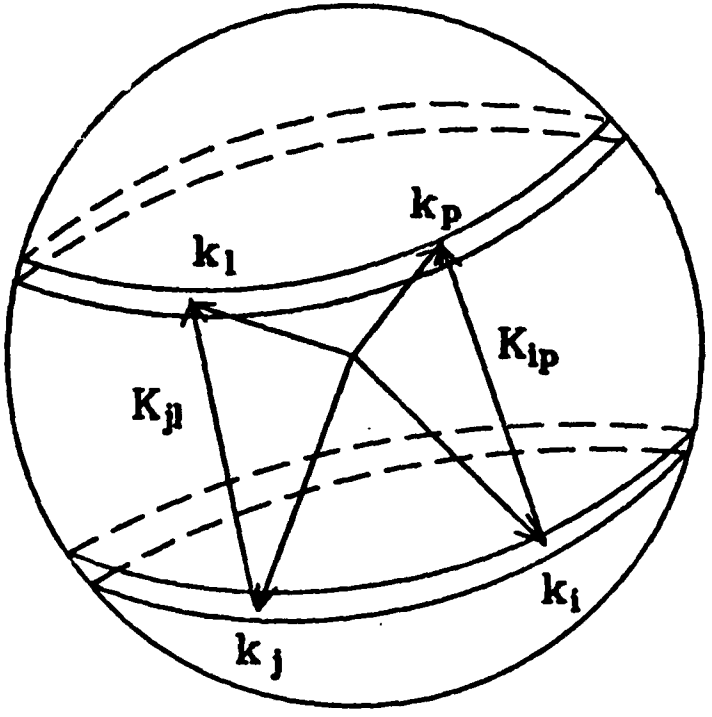
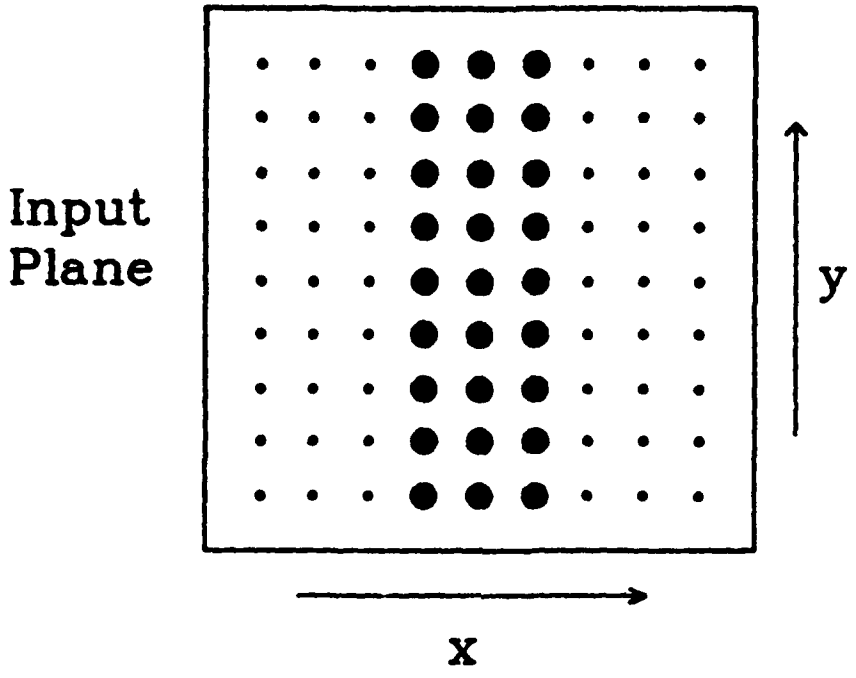


Fig. 3



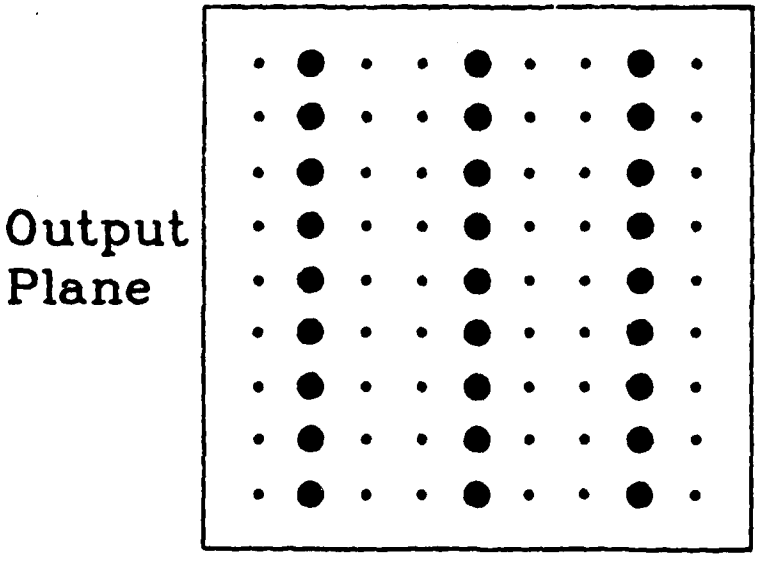


Fig. 4a

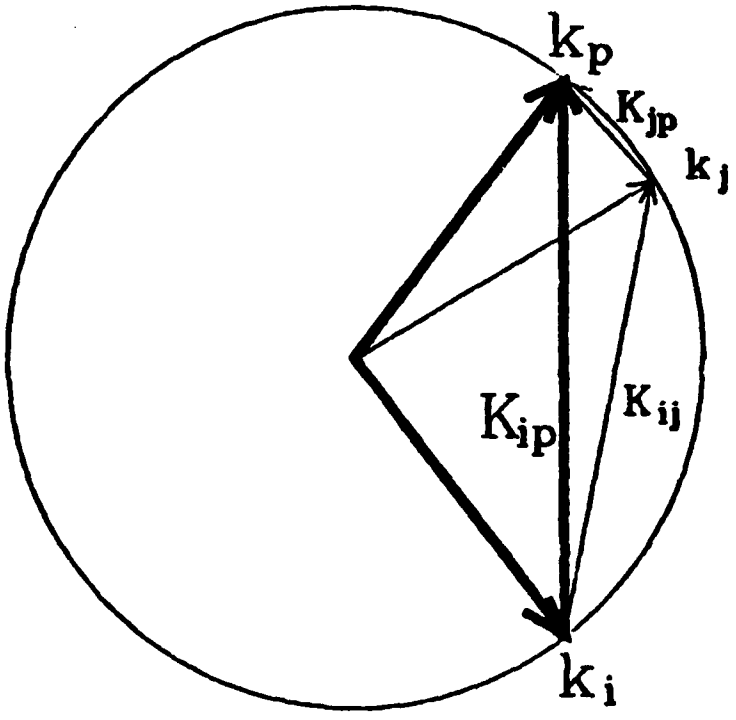
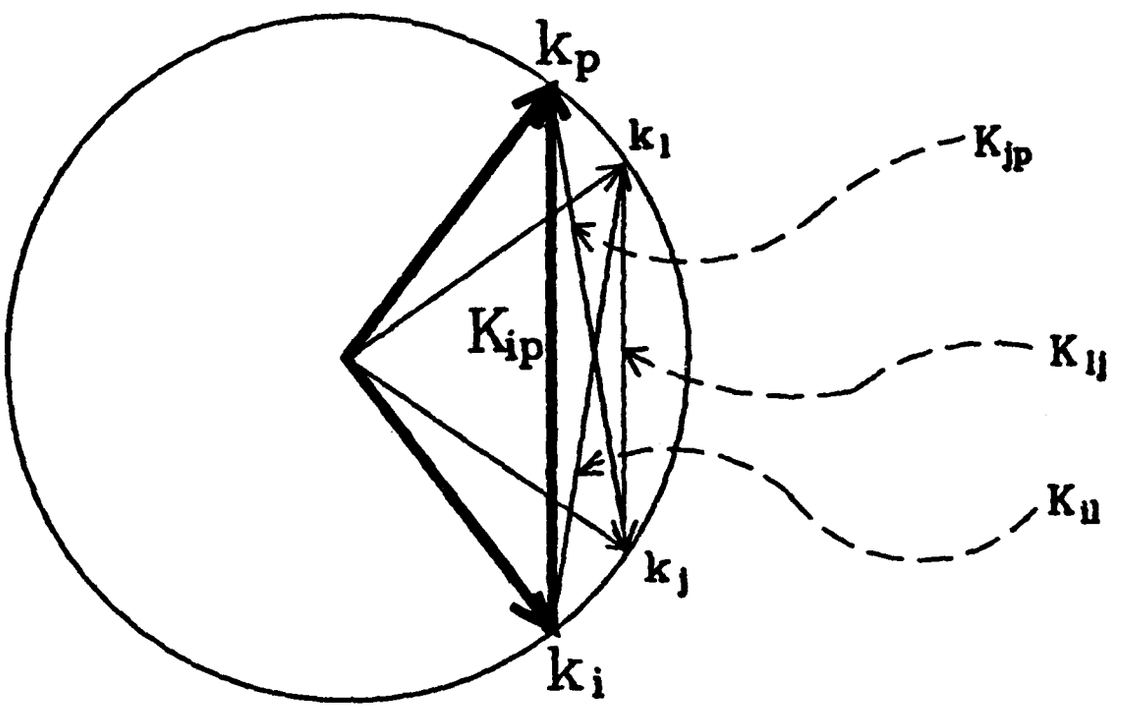


Fig. 4b



Reprinted from Applied Optics, Vol. 26, page 5061, December 1, 1987

Multilayer optical learning networks

Kelvin Wagner and Demetri Psaltis

A new approach to learning in a multilayer optical neural network based on holographically interconnected nonlinear devices is presented. The proposed network can learn the interconnections that form a distributed representation of a desired pattern transformation operation. The interconnections are formed in an adaptive and self-aligning fashion as volume holographic gratings in photorefractive crystals. Parallel arrays of globally space-integrated inner products diffracted by the interconnecting hologram illuminate arrays of nonlinear Fabry-Perot etalons for fast thresholding of the transformed patterns. A phase conjugated reference wave interferes with a backward propagating error signal to form holographic interference patterns which are time integrated in the volume of a photorefractive crystal to modify slowly and learn the appropriate self-aligning interconnections. This multilayer system performs an approximate implementation of the backpropagation learning procedure in a massively parallel high-speed nonlinear optical network.

I. Introduction

There has been considerable interest in the optics community in recent years in the optical implementation of neural network models,[1-5] and these have been considered principally for associative memory applications.[6-15] Incoherent optoelectronic implementations of matrix vector multipliers with nonlinear electrical feedback were used to demonstrate that imperfect analog hardware worked surprisingly well in the robust environment of a neural network.[7] Holographic association with coherent light can be combined with optical nonlinearities within a strongly pumped phase conjugate mirror,[8-10] or with the nonlinear thresholding capabilities of an optical spatial light modulator,[6] to implement image association. Volume holograms can be repetitively exposed to a number of Bragg angle multiplexed connectivity patterns to produce a holographic interconnection matrix.[11] These systems are programmed to perform a fixed operation by precalculating the interconnections with an easy learning procedure, so that fixed points of the idealized neural dynamics are the desired associative recall. One of the most intriguing properties of a neural network is the ability to learn dynamically the interconnections that correspond to a desired behavior through an iterative adaptation of the weight matrix through outer product perturbations.[1,4,5] Optical implementations of adaptive associative memories using optoelectronic components and spatial light modulator technology have been suggested.[12] A fascinating all-optical nonlinear dynamical system for adaptive association based on a saturating cubic nonlinearity in a phase conjugating dynamic volume holographic resonator has been proposed.[13] An even more powerful learning paradigm, sometimes called hard learning, involves either error driven learning, reinforcement learning, or self-organizing principles.[4,5] A hybrid electrooptical approach to Boltzmann learning has been proposed that is based on an incoherent optoelectronic matrix–vector multiplier interfaced with a microcomputer.[14] Error driven behavioral modification has the ability to sense system performance and adapt the synaptic weights in a manner which will compensate for some of the device imperfections and interconnection misprogrammings that caused the unwanted behavior. This paper explores the match between the backpropagation error driven multilayer learning procedure[1,2] and optical networks,[15,16] while ignoring the biological implausibility of bidirectional synapses, because of the intrinsic bidirectionality of optical interconnections. This system is a feed forward multilayer perceptron which has the potential of more general computationally universal behavior than single-layer associative networks. However, it differs from the recurrent networks because all the feedback dynamics are involved in training the modifiable interconnections and not in processing the input. We propose a new optical implementation of this multilayer learning system which uses self-aligning volume holograms to bidirectionally interconnect nonlinear etalons which act as the bidirectional optical neurons. This architecture combines the robustness of the distributed neural computation and the backpropagation learning procedure with the high speed processing of nonlinear etalons, the self-aligning ability of phase conjugate mirrors, and the massive storage capacity of volume holograms to pro-

When this work was done both authors were with California Institute of Technology, Pasadena, California 91125; K. Wagner is now at University of Arizona, Optical Sciences Center, Tucson, Arizona 85721.

Received 28 May 1987.
0003-6935/87/235061-16$02.00/0.

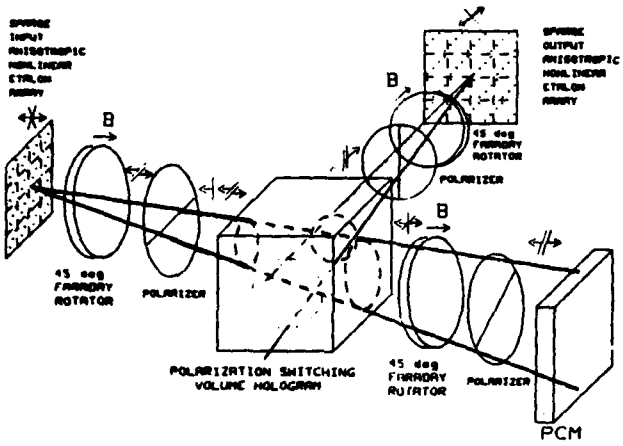


Fig. 1. Optical backward error propagation architecture with polarization multiplexed forward and backward waves, nonreciprocal polarization filtering, and self-aligning polarization switching volume hologram.

duce a powerful and flexible parallel optical processor.

One version of a single layer of this optical backpropagation architecture is shown in Fig. 1, and the operation is briefly described before discussing the idealized backpropagation algorithm and the details of this optical implementation. The learning algorithm in this single-layer optical perceptron begins with the repetitive presentation to the network input of the set of training patterns in a uniformly random sequence. Initially, the system gives rise to a sequence of output patterns through the holographic interconnection and output nonlinearity, which is different from the desired target response sequence. An error pattern is formed, either electronically or optically, by taking the difference between the actual output pattern and the targeted response. The difference pattern is sent backward through the output neurons and into the network using the same etalons and holographic interconnections, but encoding the error with an orthogonal polarization, or a slightly different frequency, or pulsed at a jittered time than the forward-propagating signal. This multiplexing of the forward and backward waves in orthogonal eigenmodes avoids direct interference between these waves. Meanwhile, the undiffracted portion of the input pattern is phase conjugated by an auxiliary phase conjugate mirror, which retroreflects each component of the input wavefront back toward the position at the input from which it originated. The phase conjugate beam has the polarization rotated or the wavelength shifted to match the error encoding to act as a self-aligning reference beam for the backward-propagating error wavefront. A volume hologram is recorded within the photorefractive crystal as the interference pattern between the phase conjugated input pattern and the backward propagating error signal. This is mathematically equivalent to changing the holographic connectivity matrix by the outer product of signal and error pattern vectors. The next time that this particular input pattern is presented to the network, it produces a diffraction pattern that more closely resembles the desired output pattern. Eventually, the hologram will learn the correspondence between a set of input patterns and the associated responses as long as the set of input patterns is linearly separable, which implies that a holographic interconnection exists that produces the desired pattern transformation. Since the holographic reference wave is generated by a phase conjugate mirror, as the network learns it will also self-align as well as correct for some of the optical imperfections present in the system components.

When the desired pattern transformation is not linearly separable, as in most difficult problems of interest, it is necessary to adaptively implement more complex nonlinear decision surfaces.[17] One way that this can be accomplished is by stacking these single-layer networks up to form a multilayer network of holographically interconnected nonlinear devices that is trainable by backpropagating the error signal through the layers. When the error pattern strikes the hologram, part of it is diffracted toward the previous layer of nonlinear devices, known as hidden units, by the transpose of the interconnection matrix seen by the forward-propagating patterns, which is the necessary connectivity for backpropagating the error. The backpropagation algorithm also requires that the transmission function of the hidden units to backward-propagating signals be the derivative of the forward mode sigmoid transfer function evaluated at the current operating level of each device. The derivative is peaked where the nonlinear sigmoid transfer characteristic has a large differential gain, so that if the hidden unit is operating in this region the connections leading to it will be strongly modified by the efficiently transmitted error signal, thereby helping that neuron to decide that it should be either high or low on subsequent presentations and not between. The multiple layers of interconnections will be continuously modified until all the patterns within the training set pro-

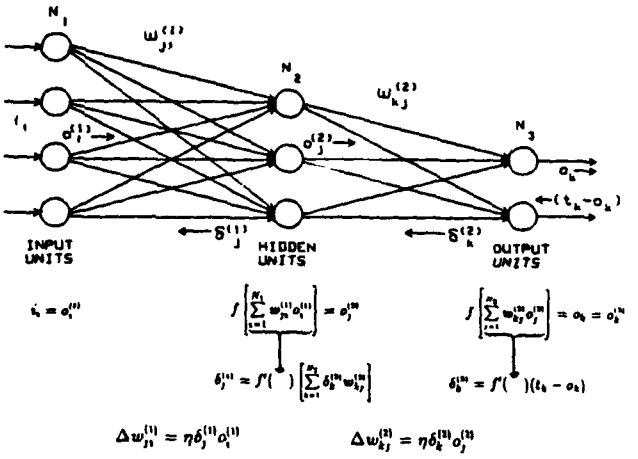


Fig. 2. Two-layer network for backpropagation learning, feed forward equations, backward-error-propagation equations, and learning rule.

duce outputs very near the flat upper or lower levels of the nonlinear device sigmoid response, so that the error signals are not allowed to backpropagate through the network. When convergence is reached, the error signals that are generated at the final layer become very small for all members of the training set.

II. Backpropagation Learning Procedure

In this section we briefly review the derivation of the backward error propagation learning procedure[1,2] to establish the notation and encapsulate the system characteristics that the optical architecture must incorporate. A schematic representation of a two-layer network is shown in Fig. 2, which consists of an input layer globally interconnected to a hidden layer, which is interconnected through a second weighted communication network to an output layer. The interconnection strengths are modifiable, so that the system can be trained to perform a desired pattern transformation from the input space to the output space. The binary signals applied to the input layer of N_1 neurons are reproduced at the output of these neurons as binary outputs, which are the inputs to the first layer, so that $i_i = o_i^{(1)}$. The outputs of the first layer are interconnected through an $N_2 \times N_1$ weight matrix $w_{ji}^{(1)}$ to a hidden layer consisting of N_2 neurons, forming presynaptic input strengths which are linear combinations of the outputs from the previous layer:

$$s_j^{(1)} = \sum_{i=1}^{N_1} w_{ji}^{(1)} o_i^{(1)}. \tag{1}$$

The hidden layer of neurons performs a soft thresholding operation on these presynaptic inputs, with a nonlinear sigmoid response $f(s)$, forming the outputs of the hidden layer which become the inputs to the second layer:

$$o_j^{(2)} = f[s_j^{(1)}] = f\left[\sum_{i=1}^{N_1} w_{ji}^{(1)} o_i^{(1)}\right]. \tag{2}$$

The outputs of the hidden layer are interconnected through the $N_3 \times N_2$ weight matrix $w_{kj}^{(2)}$, which gives the N_3 presynaptic network input to the final output layer, as a linear combination of the hidden layer outputs:

$$s_k^{(2)} = \sum_{j=1}^{N_2} w_{kj}^{(2)} o_j^{(2)}. \tag{3}$$

The final layer performs the same nonlinear soft thresholding operation as the hidden layer giving the N_3 network outputs:

$$o_k = o_k^{(3)} = f[s_k^{(2)}] = f\left[\sum_{j=1}^{N_2} w_{kj}^{(2)} o_j^{(2)}\right]. \tag{4}$$

These outputs represent the response of the network for a given set of inputs i_i, and it is the job of the training procedure to modify the interconnection weight matrices so that the actual response closely approximates the desired system response. Not all input–output mappings are possible in a network of a specific size, but complex problems of a cognitive nature with fuzzy decision boundaries have been efficiently performed in a multilayer network of this type.[18]

The desired response for the input $i_i(n)$, presented at the input of the network on the nth machine cycle, is given by a target vector $t_k(n)$, which differs from the network output $o_k(n)$, so the network error vector is given by $\delta_k(n) = [t_k(n) - o_k(n)]$. A positive definite mean-squares error (MSE) energy functional can be formed to characterize the systems behavior, and minimizing this function for all n will improve the quality of the behavior of the multilayer network:

$$E(n) = \tfrac{1}{2} \sum_{k=1}^{N_3} [t_k(n) - o_k(n)]^2. \tag{5}$$

A gradient descent procedure can be employed to modify the elements of the weight matrices and push them in the direction that improves the network performance, as measured by the MSE energy function, on subsequent presentations of a given pattern:

$$\Delta w_{kj}^{(m)} \propto -\frac{\partial E}{\partial w_{kj}^{(m)}}. \tag{6}$$

This weight update rule is designed to move the weights in a direction that rolls down the gradient of the energy surface in an amount which is proportional to the local slope. Ideally, the energy function should be averaged over the entire set of training patterns, so that the modification of the weight matrices is in the appropriate direction to improve the system response for the entire training set. However, a temporally localized learning can be performed by using a small acceleration coefficient η and modifying the weights after individual pattern presentations. The modification of the weights that results after cyclically presenting the training set in arbitrary order many times can approximate the desired change. The gradient descent is calculated by using the chain rule and representing the derivative of the energy function with respect to the weight matrix elements as a product of two parts, the backpropagating error and the forward-propagating signal:

$$\frac{\partial E}{\partial w_{kj}^{(m)}} = \frac{\partial E}{\partial s_k^{(m)}} \frac{\partial s_k^{(m)}}{\partial w_{kj}^{(m)}} = -\delta_k^{(m)} o_j^{(m)}. \tag{7}$$

The derivative of the energy with respect to the presynaptic input to the mth layer is defined to be $-\delta_k^{(m)}$, which is the backpropagating error signal in that layer. In the final layer this term is similar to the standard form of a least-mean-squares (LMS) error signal, as originally derived for the single-layer Adaline:[19]

$$\delta_k^{(2)} = -\frac{\partial E}{\partial s_k^{(2)}} = -\frac{\partial E}{\partial o_k^{(3)}} \frac{\partial o_k^{(3)}}{\partial s_k^{(2)}} = (t_k - o_k) f'[s_k^{(2)}]. \tag{8}$$

The first term is found by directly differentiating the energy function, which yields the standard error signal used in adaptive filters, and the second term is found by differentiating the nonlinear response of the

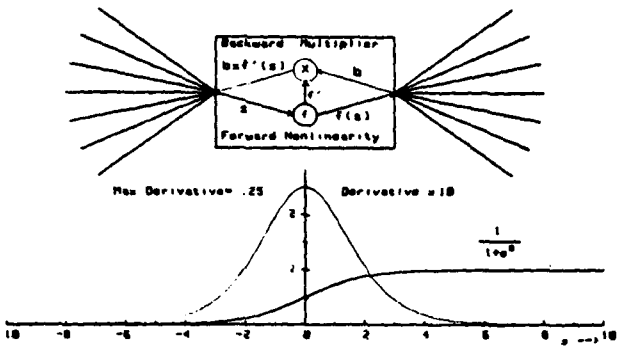


Fig. 3. Bidirectional neuron for backpropagation, its forward-mode saturating nonlinearity, and magnified derivative.

neurons. The significance of Eq. (9) is that to translate the kth component of the output error vector back into the final layer of the network it must simply be multiplied by a value which is locally computable within the kth output neuron. Thus the network output error function $\delta_k = (t_k - o_k)$ can be sent back through the corresponding output layer neurons, which multiply the error component by the derivative of the nonlinear sigmoid response at the current operating level of that output neuron. The error signal which is used to program the weights of the final layer is propagated back through those weights by multiplying by $w_{kj}^{(2)}$, and all the appropriately weighted error signals converging on the jth hidden neuron are summed to form a backpropagating presynaptic network input. The weighted sum of the error functions transmitted in the backward direction by the final layer is computed using the same interconnection matrix seen in the forward processing mode, but summing over the N_3 output neurons using the transpose of the matrix which is used for the forward-propagating interconnection:

$$\delta_j^{(1)} = -\frac{\partial E}{\partial s_j^{(1)}} = -\frac{\partial E}{\partial o_j^{(2)}}\frac{\partial o_j^{(2)}}{\partial s_j^{(1)}}$$

$$= -\left[\sum_{k=1}^{N_3}\frac{\partial E}{\partial s_k^{(2)}}\frac{\partial s_k^{(2)}}{\partial o_j^{(2)}}\right]\frac{\partial o_j^{(2)}}{\partial s_j^{(1)}} = \left[\sum_{k=1}^{N_3}\delta_k^{(2)}w_{kj}^{(2)}\right]f'[s_j^{(1)}]. \quad (9)$$

This represents an iterative algorithm for successively computing the error function at deeper layers back toward the beginning of the network in terms of the error function injected back into the final layer. Alternatively, this algorithm can be considered to represent a wavefront that backpropagates through the network, multiplying by the weights, accumulating at the neurons, and multiplying by the neurons backward transmittance to compute the appropriate error to program the previous layer. The network is highly nonlinear in the forward-propagating direction, but the backpropagating wavefront is computed using only linear operations.

The neurons must, therefore, have two signal pathways as shown in Fig. 3. The two pathways share the same weights on the connected layers, but the neuron response is a nonlinear soft thresholding for forward-propagating signals and a multiplier that only allows a backward propagating error through the neuron when the slope of the forward mode operation is large. The transmitted components of the backpropagating error vector are only large when the corresponding output neurons are operating in the steep thresholding regime where the derivative is large, and that component was significantly in error at the network output. Any neuron that had decided that it is a one or zero by being well above or below the threshold knee inefficiently transmits the error back into the previous layer of the network. From the definition of the change in the weight matrix given in Eq. (6) and the chain rule expansion of Eq. (7) we can write the form of the weight update rule for the mth layer according to this first-order gradient descent procedure:

$$w_{kj}^{(m)}(n+1) = w_{kj}^{(m)}(n) + \eta\delta_k^{(m)}o_j^{(m)}. \quad (10)$$

The error transmitted by the neurons back into the previous layer of interconnections is used to modify the weights of that layer through this outer product update rule. The weighted interconnection $w_{kj}^{(m)}$ is carrying the output from the jth neuron in the mth layer $o_j^{(m)}$ to the kth neuron in the $(m+1)$st layer, which is simultaneously broadcasting the error function $\delta_k^{(m)}$ back into the mth layer of the network. The product of this forward-propagating signal and backward-propagating error takes place within each weighted synaptic connection as the desired weight update contribution, completely independently of what is taking place within all the other weighted connections, and this is the only information needed to update that weight, so this learning rule can be said to be a local update rule. The training procedure for the final layer of weights is given by an appropriate outer product learning rule, which is a local update rule that takes place within each weighted signal pathway, but the problem of credit assignment of the MSE energy to the earlier layer has been solved by nonlocally backpropagating the error vector. This is referred to as the backward error propagation algorithm for training multilayer networks, and it can be further generalized to N layer networks or networks with feed forward interconnections, e.g., when the first layer connects directly with the output layer as well as indirectly through the hidden layer to the output layer. For more details of the derivation, operation and utility of this multilayer network training algorithm the reader is referred to Refs. 1, 2, and 18.

III. Optical Implementation

The optical implementation of a backpropagation network requires two basic bidirectional components, the interconnection matrices, and the nonlinear units. Volume holograms appear as the most promising candidate for implementation of an interconnection matrix because of the large storage capacity possible within the volume of a crystal and the dynamic response possible with a photorefractive crystal. The readout

of a volume hologram can be accomplished with either a forward-propagating beam or a backward-propagation. Spatial light modulators (SLMs) could also be used as the interconnection element for small networks, but they would have to be both bidirectional and optically addressable to be used in a backpropagation network. In this paper a new self-aligning approach to adaptively forming optical interconnections based on phase conjugating one of the undiffracted beams is presented. This technique uses interferometric detection in the volume of a photorefractive crystal to accomplish all the outer product multiplications necessary for weight matrix perturbation. These $N_1 \times N_2$ weight updates are calculated in parallel by exposing the crystal with N_1 phase conjugated collapsing spherical waves and N_2 expanding spherical waves simultaneously.

The nonlinear units or neurons need to threshold the forward-propagating beam while transmitting the backward beam only when the forward beam nonlinearity is in the high slope, or undecided, regime of operation. A special purpose, bidirectional, detector modulator pair array structure could be tailored to generate the desired backpropagation neuron responses by utilizing the appropriate integrated electronic circuitry, but the individual neurons could become quite complicated with this conventional optoelectronic integrated circuit approach. Appropriately modified transmissive spatial light modulators might be considered for backpropagation neurons, and one possible structure of this type is illustrated in Fig. 4. In this type of birefringent SLM, crossed polarizers are placed on either side of an electrooptic medium which is optically addressed by a photoconductor. A high voltage is applied across a transparent conductor in contact with the photoconductor on one side and a transparent conductor on the other. To use this type of electrooptic device as a backpropagation neuron the induced birefringence must be doubled for the backward propagating wave to obtain a saturating forward nonlinearity while obtaining a derivative backward multiplication. This can be accomplished by a pair of photoconductors, both addressed by the same forward propagating beam, where one is used to modulate the forward-propagating device which is biased with a voltage V_π, while the other is used to modulate an EO device with a saturation voltage $2V_\pi$. The forward-propagating modulator is used to modulate a fixed intensity pump I_0 so that a single half-cycle of a saturating nonlinearity can be generated, $I_f^m = I_0 \sin^2(I_f^{m-1}/I_{sat})$ for $I_f^{m-1} < I_{sat}$, and $I_f^m = I_0$ otherwise. The backward-propagating modulator is used to multiply the backward-propagating error signal by a function $I_b^{m-1} = I_b^m \sin^2(2I_f^{m-1}/I_{sat})$ for $I_f^{m-1} < I_{sat}$, and $I_b^{m-1} = 0$ otherwise, and this is of the form of the desired derivative multiplication. Since the two functions required of a backpropagation neuron can also be accomplished with a simpler nonlinear resonator structure, and the response time of these nonlinear etalons can be extremely fast compared to SLM technology, they were chosen for study in the architecture presented in this paper.

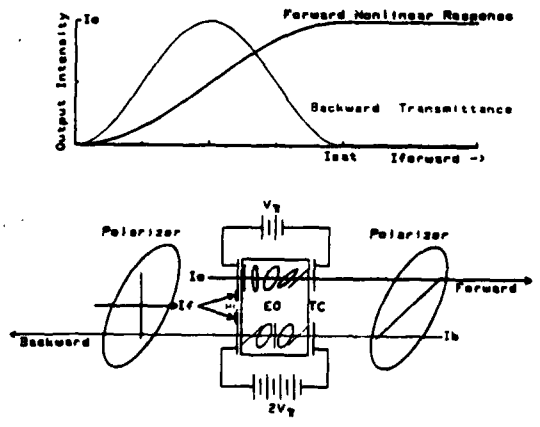


Fig. 4. Input output relations for a special purpose bidirectional optically addressed spatial light modulator backpropagation neuron: PC = photoconductor; EO = electrooptic; TC = transparent conductor.

A. Nonlinear Fabry-Perot Backpropagation Neurons

Nonlinear Fabry-Perot etalons[20] are a promising candidate for implementing the neurons in an optical learning network because they can perform nonlinear operations on arrays of coherent beams, which allows the outputs to be used to record and modify interconnection holograms. A soft thresholding operation can be performed on a forward-propagating beam by decreasing the cavity detuning below the critical detuning needed for bistability.[20] These optical neurons cannot easily implement the idealized derivative transmission required for backpropagation, but a similar peaked response can be obtained by operating a nonlinear etalon in the probe mode[21] for the backward-propagating error signal. In this mode, the Fabry-Perot resonance is scanned by the nonlinear dependence of the index on the intracavity intensity, which varies in response to the high power forward beam intensity. The weak backward-propagating probe beam does not scan the cavity, but it is modulated by the current state of the cavity tranamission function, which is the appropriate multiplication type of response needed in the backward direction. The probe mode transmission is peaked at the resonance of the Fabry-Perot, which occurs when the sigmoid response to the forward beam reaches the upper level. The peak maximum is not exactly at the region of the highest slope of the forward beams nonlinear sigmoid response, but since the forward and backward beams have different polarizations, or different wavelengths, the resonance function can be offset to achieve a properly positioned probe beam resonance peak.

In the polarization multiplexed case this shift can be induced by including a thin birefringent sheet in the cavity,[22] or perhaps a tunable birefringence can be caused by applying a static external field to the cavity. This type of birefringent nonlinear Fabry-Perot etalon

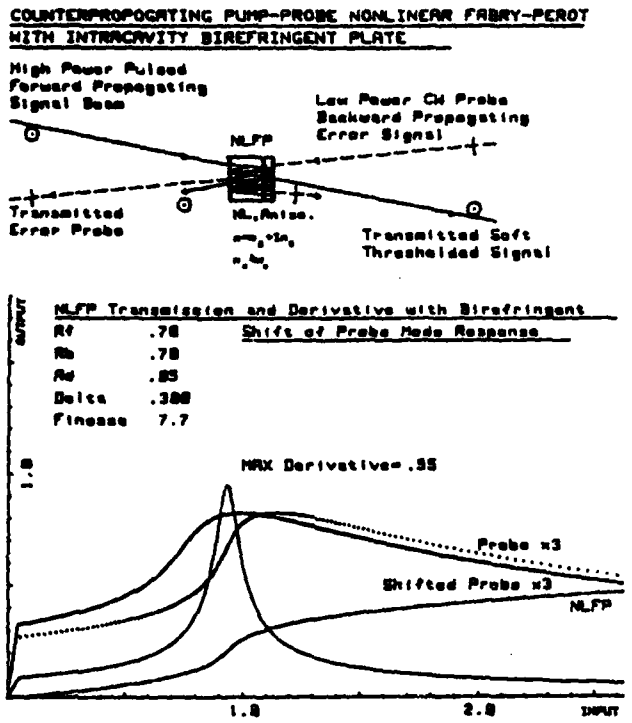


Fig. 5. Nonlinear Fabry-Perot etalon sigmoid response, its derivative, and the probe mode transmission for the two polarizations with an auxiliary intracavity birefringence.

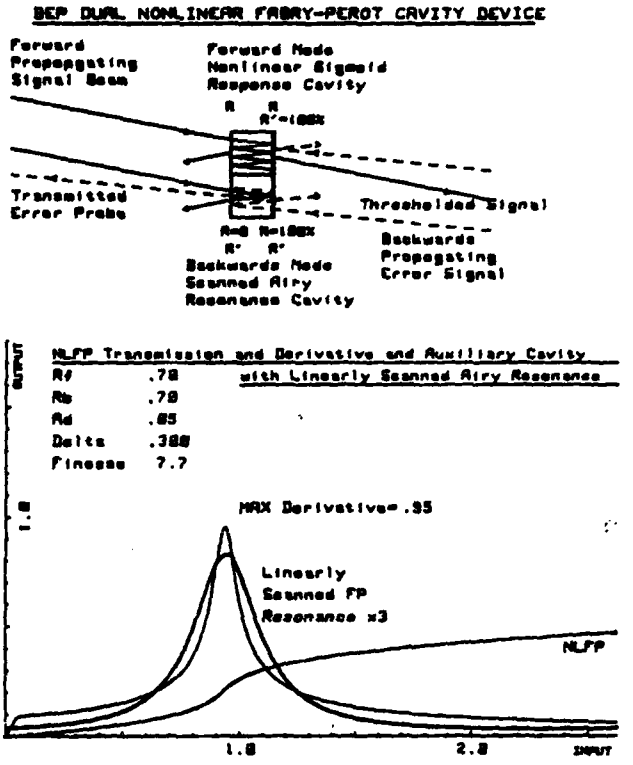


Fig. 6. Dual-cavity nonlinear Fabry-Perot etalon with forward-propagating nonlinear response and backward-propagating scanned resonance probe mode transmission

and a simulation of the forward-mode sigmoid transfer function is shown in Fig. 5 along with its derivative and the shifted probe mode response approximation to this derivative. This device can implement the desired sigmoid nonlinearity of the high intensity forward-propagating signals with a differential gain greater than one, although the actual gain in transmission is less than one. The probe mode response is not symmetric about the peak because the Airy function resonance is scanned by the intracavity intensity which is equal to the transmitted sigmoid response divided by the backmirror transmittance. This asymmetry continues to allow signals that are above threshold to build up interconnection gratings in the previous stage corresponding to correlated inputs, thereby partially compensating for the slow forgetting of gratings by the volume hologram. However, the high level of transmission for the probe beam when the etalon pump is below threshold is undesirable. By decreasing the finesse of the cavity to the forward-propagating beam a trade-off can be made between the peak width and off-resonance transmission of the probe mode response, with the switching energy for the forward-propagating nonlinear device characteristic.

Another possibility would be to use two closely spaced cavities, both addressed by the same forward- and backward-propagating resolution spots, as illustrated in Fig. 6. In this case one cavity is optimized to produce a sigmoid response of the forward-propagating beam while blocking the backward-propagating error signal, while the other cavity is resonant to the backward- propagating beam. The Fabry-Perot resonance of the backpropagating cavity is linearly scanned by the 100% reflected forward-propagating incident intensity, thereby producing a good approximation to the desired symmetric derivative response. We expect that learning and eventual convergence can be achieved in a multilayer optical network with the forward and backward response that can be obtained from these scanned resonance devices, even though the responses do not precisely match the nominal responses of the backpropagation algorithm because of the robustness of this gradient descent learning procedure.

B. Description of a Single Layer of the Optical Architecture

A single layer of an architecture that can perform this type of multilayer perceptron learning procedure, using polarization multiplexing of the forward-propagating processing beam and backward-propagating teaching beam, is shown in Fig. 1. The illustrated architecture is one implementation of this class of backward-error-propagating holographic learning machines that serves to illustrate the principles involved. Notice that no lenses are shown in this diagram because the volume hologram can perform the desired weighted interconnection imaging by exposing it with

the proper expanding image and focusing reference beam to form a Fresnel volume hologram. If Fourier lenses are inserted between the etalon arrays and the volume holographic crystal, the exposed hologram will be a Fourier hologram with planar fringes, and the momentum space analysis will be simplified, but the processor learning and self-aligning operations will be similar.

The forward-propagating pattern vector transmitted by the anisotropic nonlinear etalon array on the left-hand side of the figure is polarized at a −45° angle and is rotated clockwise by 45° as it passes through the nonreciprocal Faraday rotator so that it becomes horizontally polarized. This aligns the forward-propagating beam with the polarizer allowing it to pass and illuminate the polarization switching volume hologram. The diffracted beam consists of a weighted interconnection of the forward-propagating pattern vector by the current state of the holographically represented interconnection matrix, stored as a superposition of curved and chirping space charge gratings within the photorefractive crystal. The diffracted beam is polarization switched by the birefringent diffraction mechanism to an orthogonal polarization to the input, and this vertical polarization state is rotated clockwise by 45° through the following Faraday rotator so that it falls on the next etalon array with the same −45° polarization as the forward-propagating transmitted beam from the previous stage. The undiffracted beam passes straight through the volume hologram and has its vertical polarization rotated by 45° as it passes through the Faraday rotator, so that it falls on the phase conjugate mirror (PCM) with a polarization angle of 45°, the same as the counterpropagating pump beams (which are not shown), producing an identically polarized phase conjugate beam, which is composed of an array of beams that retroreflect back toward the etalon sources that each originated from. This phase conjugate beam passes through the nonreciprocal Faraday rotator picking up another 45° rotation (instead of unwrapping the rotation as would occur with a reciprocal optical activity based rotator), emerging vertically polarized to act as the reference beam array for the self-aligning holographic outer product exposure with the backward-propagating error signal. The backward-propagating error signal emerges from the backside of the output etalon array with a 45° polarization that is orthogonal to the forward-propagating beam. This allows for the polarization filtering based separation of the reflected forward-propagating beam from the transmitted backward-propagating beam as well as the independent tuning of the relative Fabry-Perot resonance position of the forward- and backward-propagating beams. The backward propagating error signal is rotated to a vertical polarization by the Faraday rotator so that it interferes in the volume hologram with the vertically polarized phase conjugate reference beam and not with the horizontally polarized undiffracted forward-propagating signal. The interference of a backward-propagating error signal emerging from a particular etalon at the output with the phase conjugated forward-propagating beam emerging from a particular etalon from the input produces a self-aligning volume Fresnel holographic interference pattern that interconnects these two etalons for both forward- and backward-propagating beams with the exact same diffraction efficiency, or weight, due to the reciprocity of linear electromagnetic systems. The interference of the backward-propagating error beam with the phase conjugate of the forward-propagating beam records a Fresnel grating due to each pair of beams that is present, perturbing the weighted interconnection matrix represented by the hologram by the outer product of the signal and error vectors and thereby pushing the matrix toward the desired interconnection solution. The backward-propagating beam is polarization switched by the volume holographic diffraction mechanism, producing a horizontally polarized beam which is the appropriate weighted summation of the error signal by the transpose of the interconnection matrix seen by the forward-propagating beam. This passes through the polarizer and is Faraday rotated by 45° to be incident on the etalon array with a 45° polarization angle, orthogonal to the forward-propagating beam, and the same as the backward-propagating beam which emerged from the previous output layer. The undiffracted phase conjugate of the forward-propagating beam needs to be blocked so that it is not confused with the copropagating diffracted backward-error-propagating signal, and this is accomplished by the polarizer which blocks the vertical polarization of the undesired phase conjugate reference beam. The indicated nonreciprocal polarization filtering will also remove the unwanted reflections from the nonlinear etalons and unwanted diffraction terms produced by the hologram. The diffracted phase conjugate reference and undiffracted backward-error signal emerge at a different angle and will not focus on the etalon; thus they can be ignored, or they can be examined to determine intermediate states of the hidden neurons. Each layer is completely compatible with the previous and the following layers so this type of learning network can be stacked up to form a complex multilayer learning machine.

C. Requirements for the Holographic Interconnection

The dynamic holographic interconnection technique described in this paper is based on the photorefractive effect, which is a light-induced index of refraction modulation that occurs in photoconductive electrooptic crystals. A space charge grating image of an interference profile is created by carriers ionized from fixed traps into the conduction band, where the mobile carriers redistribute under the influence of drift, diffusion, and bulk photovoltaic effects, until they recombine with an unoccupied trap. The redistributed optically generated carriers produce a space charge grating with a fundamental Fourier component that may be phase shifted from the interference profile. The spatial variations of the resulting space charge pattern produce a corresponding electric field through Poisson's equation. This space charge field

induces an electrooptic modulation of the local impermeability tensor as long as the appropriate electrooptic tensor coefficient is nonzero. In turn, this couples the input field into a diffracted output field as long as the appropriate impermeability tensor coefficient is nonzero. The implementation presented in the previous section is based on a polarization switching diffraction mechanism for which it is required to have an off-diagonal impermeability tensor coefficient. This requires electrooptic tensor coefficients in the bottom half of the reduced subscript electroopotic matrix, which can take place in some electrooptic volume holographic materials, such as $Bi_{12}SiO_{20}$, $LiNbO_3$, $BaTiO_3$, and GaAs. Self-aligning recording combined with polarization switching diffraction between linear eigenmodes requires an optically isotropic medium (or one in which anisotropy can be eliminated through the application of a static field), with no optical activity, and these conditions imply that a material of symmetry group $\bar{4}3m$, such as some III–V semiconductors (e.g., GaAs or InP), should be used as the photorefractive holographic medium. The efficiency of the diffraction depends on the effective coupling strength which depends on the angle of the gratings, polarization of the input wave, and momentum matching (Bragg) condition in a rather complicated fashion. However, use of a Fresnel hologram can produce an averaging over all these effects for all the interconnection holograms, while in a Fourier hologram with planar fringes each grating has a different diffraction efficiency.

The polarization switching diffraction efficiency and the holographic storage capacity can be simultaneously maximized by having the input and output beams propagating at large angles, as indicated in the figure. The unwanted polarization switching grating exposures due to the simultaneous presence of multiple reference (or object) beams produce crosstalk of the undiffracted forward-propagating beam, which can be eliminated with the indicated polarization filtering. The storage capacity of the volume hologram will enforce limits on the number of nonlinear devices that can be interconnected and on their topology because of the cone of ambiguity associated with Bragg diffraction.[23] A sparse array of etalons will have to be utilized to implement a fully global interconnection without unwanted crosstalk, which will also facilitate the dissipation of heat generated in the nonlinear etalons and thereby allow a very high speed of operation. However, the learning operation must occur slowly for the backpropagation algorithm to converge properly, and this is well matched with most photorefractive crystal volume holograms, because the crystal response times are slow, and the perturbation of an existing space-charge grating by a single outer product exposure is very small.

It is necessary to be able to both selectively erase holographic gratings, thus decreasing the connection strength between particular etalons, as well as to strengthen individual gratings thereby increasing the corresponding elements of the interconnection matrix. Selective erasure can be accomplished by using a phase encoded backward-propagating error signal, where a phase angle of 0 is used to represent all positive error signals, and a phase angle of π is used to represent all negative error signals. Fresnel gratings that are built up with a phase angle of 0 can have the corresponding interconnection decreased selectively by shifting the recording interference profile by π, as demonstrated in Sec. V, and by Huignard for Fourier holograms.[24] Alternatively, selective interconnection erasure might be accomplished by strengthening interconnection gratings when the applied bias field is in one direction, causing the resulting space charge grating to shift away from the optical intensity profile in the direction of the E field by approximately $\pi/2$, while decreasing interconnection gratings when the bias field is reversed, producing a canceling space charger grating with a phase shift of $-\pi/2$. Another approach to decreasing interconnection strength would be to rely on the simultaneous erasure of all the gratings by the optical readout and thermal effects, thereby inserting a forgetting term in the dynamical equation for the holographically represented interconnection matrix. This approach requires continuous reinforcement to avoid forgetting everything that has been learned. Once learning has been completed a mechanism of fixing the hologram could be used to make the interconnections permanent.[25]

A scheme must be devised to implement negative interconnection stengths, or else all the signals must be placed on an appropriate bias. An attractive possibility for the implementation of bipolar weights is to use the phase shift of each grating to represent its sign and count on destructive interference within each nonlinear etalon to subtract the positively and negatively weighted diffracted components. This approach is sensitive to the phase response of the etalons, so it is necessary to minimize (or to compensate for) nonlinear phase shifts produced by the etalons and to avoid phase sensitive switching behavior in the etalons.[26]

IV. Self-Aligning Bidirectional Volume Holographic Interconnections

The preliminary analysis of a bidirectional optical interconnection system begins with an explanation of the recording of a hologram by using a phase conjugated reference beam. The 1-D system used in this analysis is presented in Fig. 7 and consists of two lines (planes) of optical neurons which need to be interconnected by a volume hologram. The phase conjugate mirror is used to conjugate the expanding waves emitted from one plane of neurons and retroreflect them back toward the sources from which they emerged, and either direction can be the one chosen to be conjugated. The field emitted by a line of J neurons, separated by D, with an aperture profile $h(x_0,y_0)$, and propagating at an off-axis angle with spatial frequency α is a linear combination of off-axis spherically expanding waves propagating toward the right. The undiffracted field that passes straight through the volume hologram (in the undepleted pump approximation) and

strikes the phase conjugate mirror (PCM) is simple in the Fraunhofer regime of the individual apertures, which for an aperture profile width of d is valid for $z \gg \pi d^2/\lambda$, ($z \gg$ 1 mm for a 10-μm aperture) and is always valid for Gaussian apertures:

$$A(x',y',z_s,t) = \exp(-i2\pi\nu t)\frac{\exp\left(i\frac{2\pi}{\lambda}z_s\right)}{i\lambda z_s}\int\int\sum_j a_j h(x_0 - jD, y_0)\exp(i2\pi\alpha x_0)\exp\{i(\pi/\lambda z_s)[(x_0 - x')^2 + (y_0 - y')^2]\}dx_0 dy_0$$

$$\approx \exp(-i2\pi\nu t)\frac{\exp\left(i\frac{2\pi}{\lambda}z_s\right)}{i\lambda z_s}\sum_j a_j \exp(i2\pi jD\alpha)\exp\{i(\pi/\lambda z_s)[(x' - jD)^2 + y'^2]\}H\left(\frac{x' - jD}{\lambda z_s} + \alpha, \frac{y'}{\lambda z_s}\right). \quad (11)$$

In this expression the neural activity pattern vector a_j is represented as a spatially multiplexed array of etalon output fields which will be used for learning, but the intensities $|a_j|^2$ are a more likely representation of the neuron outputs that will be used for subsequent nonlinear processing.

Each source produces an off-axis expanding spherical wave with a linear phase factor given by the source position. The Fourier transform of a source aperture is given by $H(u,v)$ and its size and position shift with the propagation distance z_s. The distance between the input plane of neurons and the front size of the volume hologram is z_0, the thickness of the hologram is L, the index of refraction of the photorefractive medium is n_0, and the distance from the hologram to the phase conjugate mirror is z_c; thus the total optical path length between the neurons and the PCM is $z_s = z_0 + n_0 L + z_c$. This wavefront is phase conjugated by the PCM, which retroreflects each expanding spherical wavefront back toward its point of origin. The resulting field within the holographic crystal is dependent on the z coordinate, and since both writing waves are incident on the hologram from the right, z is defined to be zero at the right edge of the crystal and increased to L at the left edge. The phase conjugated reference wave within the hologram is most easily expressed in terms of the optical path length between the input neurons and a given plane within the hologram, $z' = z_0 + (L - z)n_0$.

$$A^*(x,y,z,t) = \exp(-i2\pi\nu t)\frac{\exp\left(-i\frac{2\pi}{\lambda}z'\right)}{-i\lambda z'}\sum_j a_j^* \exp(-i2\pi jD\alpha)$$
$$\times \exp\{-i(\pi/\lambda z')[(x - jD)^2 + y^2]\}$$
$$\times H^*\left(\frac{x - jD}{\lambda z'} + \alpha, \frac{y}{\lambda z'}\right). \quad (12)$$

This equation represents a left propagating quadratically curved superposition of waves that are focusing toward the J source neurons. The profiles of these focusing beams are given by the transform of the individual source apertures $H(u + \alpha', v)$ that are scaled and shifted with the z coordinate within the holographic medium.

Similarly, the backpropagating error field emitted by a line of K neurons separated by D' in the second layer can be described as a superimposition of spherically expanding waves. The separation between the output neurons and a plane z within the hologram is $\bar{z} = z_1 + n_0 z$, which is a reversed coordinate from that used for the forward-propagating wave:

$$B(x,y,z,t) = \exp(-i2\pi\nu t)\frac{\exp\left(i\frac{2\pi}{\lambda}\bar{z}\right)}{i\lambda\bar{z}}\int\int\sum_k b_k h(x_1 - kD', y_1)\exp(i2\pi\alpha x_1)\exp\{i(\pi/\lambda\bar{z})[(x_1 - x)^2 + (y_1 - y)^2]\}dx_1 dy_1$$

$$\approx \exp(-i2\pi\nu t)\frac{\exp\left(i\frac{2\pi}{\lambda}\bar{z}\right)}{i\lambda\bar{z}}\sum_k b_k \exp(i2\pi kD'\alpha)\exp\{i(\pi/\lambda\bar{z})[(x - kD')^2 + y^2]\}H\left(\frac{x - kD'}{\lambda\bar{z}} + \alpha, \frac{y}{\lambda\bar{z}}\right). \quad (13)$$

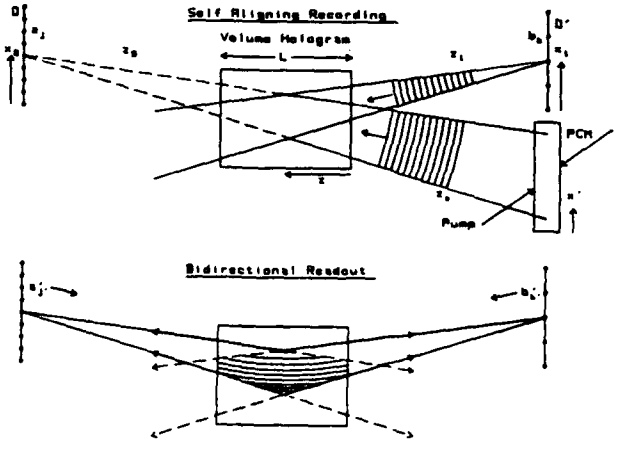
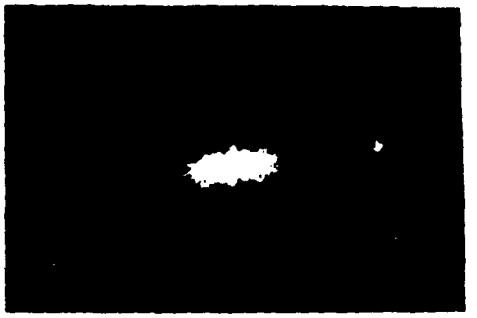
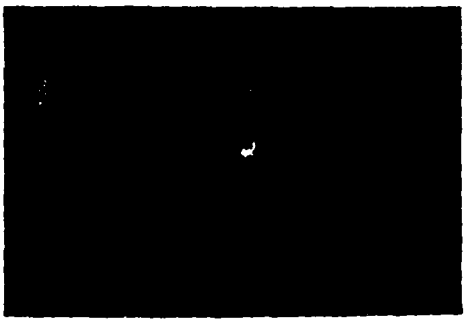


Fig. 7. Self-aligning bidirectional dynamic volume holographic interconnection using a phase conjugated reference



(a)



(b)

Fig. 8. Diffracted spot produced by a high diffraction efficiency lensless Fresnel volume hologram recorded in $LiNbO_3$ and an exposure which shows just the peak.

This backpropagating error vector $\mathbf{b}_k$ produces a wavefront that interferes with the retroreflected phase conjugate wavefronts due to the forward-propagating sources in the first layer within the volume hologram. This records a self-aligning interference pattern that modulates the index of refraction within the holographic medium. The modulation term can be expressed as a curved and chirping fringe pattern within the overlap region of the diffracting wavefronts in the crystal. The repetitive presentation of training patterns and bipolar error patterns to the front and back of the single layer being described will result in the time integration of successive outer product connectivity patterns:

$$T(x,y,z,t) \propto \int_0^t \mathcal{R}[A^*(x,y,z,t')B^*(x,y,z,t')]dt'$$

$$= \sum_j \sum_k \left[\int_0^t a_j^*(t')b_k^*(t')dt'\right] \frac{1}{\lambda^2 z' z} \mathcal{R}\left[H^*\left(\frac{x-kD'}{\lambda z}+\alpha,\frac{y}{\lambda z}\right)H^*\left(\frac{x-jD}{\lambda z'}+\alpha,\frac{y}{\lambda z'}\right)\right]$$

$$\times 2\cos\left(\frac{2\pi}{\lambda}\left\{z_t+\frac{[(x-jD)^2+y^2]}{z'}+\frac{(x-kD')^2+y^2}{z}+\lambda(jD+kD')\right\}\right). \quad (14)$$

This expression represents a superposition of KJ families of elliptical fringes within the volume of the hologram with each pair of sources at the foci of a set of elliptical shells, and $z_t = z_0 + n_0 L + z_1$ is the total distance between the etalon planes. This time-integrated interference pattern will be transformed into a proportional index modulation with a possible phase shift through the photorefractive effect.

For the chirped and curved volume Fresnel phase holograms being analyzed here a momentum space analysis is inappropriate, since spatial frequency and fringe tilt are spatially varying, resulting in a poorly defined perturbation momentum vector; thus an explicit integration of the diffracted field produced at each z-plane should be carried out instead. After the hologram is recorded, it is reilluminated by a weighted superposition of expanding spherical waves which are diffracted by all the index modulations that are present. This analysis can be carried out for either forward- or backward-propagating waves in an identical manner, but we only consider illumination with a forward-propagating wave here. When the volume hologram is illuminated by the diffracted wavefront from a new input neural activity pattern $a'_{j'}$, the diffracted field at each plane z will contain a matched term, which will produce focusing wavefronts propagating toward the output neurons and a number of

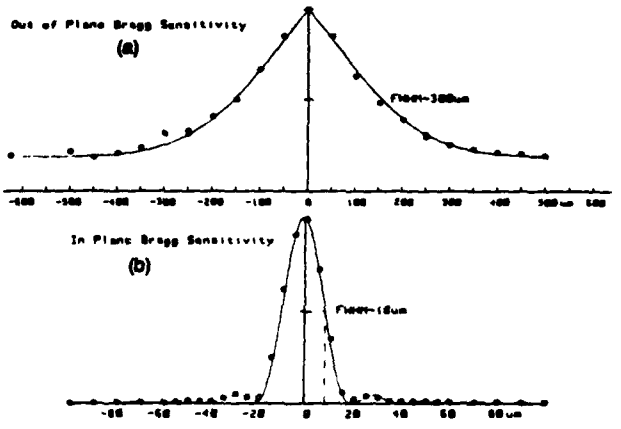


Fig. 9. Positional sensitivity of the Fresnel volume hologram: (a) out of the primary interaction plane; (b) in the interaction plane.

unwanted crosstalk terms. Combining the results of Eqs. (12) and (13) we can derive an expression for the diffracted field produced at each plane of the volume hologram:

$$D(x,y,z,t) \propto A'(x,y,z,t)\int_0^t \Re[A^*(x,y,z,t')B^*(x,y,z,t')]dt'$$

$$\propto T(x,y,z)\exp(-i2\pi\nu t)\frac{\exp\left(i\frac{2\pi}{\lambda}z'\right)}{i\lambda z'}\sum_{j'}a_{j'}'\exp(i2\pi j'D\alpha)\exp[i(\pi/\lambda z')[(x-j'D)^2+y^2]]H\left(\frac{x-j'D}{\lambda z'}+\alpha,\frac{y}{\lambda z'}\right)$$

$$=\exp(-i2\pi\nu t)\frac{1}{\lambda^2 z'^2}\sum_{j'}a_{j'}'\sum_j\sum_k\int_0^t a_j^*(t')b_k^*(t')dt'$$

$$\times\frac{\exp\left(-i\frac{2\pi}{\lambda}\bar{z}\right)}{-i\lambda\bar{z}}\exp[i2\pi\alpha[(j'-j)D+kD']]\exp[i(\pi/\lambda z')[(j'^2-j^2)D^2-2Dx(j-j')]]\exp[-i(\pi/\lambda\bar{z})[(x-kD')^2+y^2]]$$

$$\times H^*\left(\frac{x-kD}{\lambda\bar{z}}+\alpha,\frac{y}{\lambda\bar{z}}\right)H\left(\frac{x-j'D}{\lambda z'}+\alpha,\frac{y}{\lambda z'}\right)H^*\left(\frac{x-jD}{\lambda z'}+\alpha,\frac{y}{\lambda z'}\right). \quad (15)$$

With the approximations that the last two terms approximately overlap so that the product is equal to a constant, and the paraxial approximation, the diffracted field can be propagated to the plane of output neurons from any diffracting plane within the volume hologram. This is assuming the undepleted pump approximation, which is a reasonable approximation for diffraction efficiency ≤10%. Each plane of infinitesimal thickness produces an appropriate focusing contribution, with the appropriate focal length, magnification, and phase to produce a focal spot with profile $h(x_1 - kD' - (j-j')D(\bar{z}/z'),y_1)$. We need to sum up all the contribution throughout the thickness L of the hologram to obtain a Bragg selection condition, which will require $j = j'$, so that focal spots are only produced at each of the K output neurons:

In this equation a number of simplifications have been made, but the neglected terms will lead to an increase in the Bragg selectivity. All the phase factors have been lumped into the term $\exp(i\varphi)$. The integration over z produces a sinc function of $(j - j')$, which is analogous to the thick hologram Bragg condition for these elliptical fringes. As long as the separations between the input neurons D, and output neurons D', are large enough for a given hologram thickness L and recording geometry, z_0,z_1,α, we can assume perfect Bragg selectivity, and, therefore, $j = j'$. This would be the normal Bragg condition if Fourier lenses were inserted in the processor, and its results in a positional selectivity in the plane of the lines of neurons which effectively eliminates all the unwanted shift-invariant crosstalk terms that are present with a thin holographic grating.

$$e(x_1,y_1,t) \propto \int_0^L\left[\frac{\exp\left(i\frac{2\pi}{\lambda}\bar{z}\right)}{i\lambda\bar{z}}\int\int D(x,y,z,t)\exp i(\pi/\lambda\bar{z})[(x_1-x)^2+(y_1-y)^2]]dxdy\right]dz$$

$$\propto \exp(-i2\pi\nu t)\exp(-i2\pi\alpha x_1)\exp(i\varphi_0)\sum_{j'}a_{j'}'\sum_j\sum_k\int_0^t a_j^*(t')b_k^*(t')dt'$$

$$\times\int_0^L h\left(x_1-kD'-(j-j')D\frac{\bar{z}}{z'}\right)\exp\left[-i\alpha D(j-j')\frac{z}{z'}\right]dz$$

$$\approx \exp(-i2\pi\nu t)\exp(-i2\pi\alpha x_1)\exp(i\varphi')\sum_{j'}a_{j'}'\sum_j\sum_k\int_0^t a_j^*(t')b_k^*(t')dt'$$

$$\times h\left(x_1-kD'-(j-j')D\frac{z_1}{z_0+nL}\right)L\,\mathrm{sinc}\left[\frac{nLaz_1D(j-j')}{(z_0+Ln)^2}\right]$$

$$\propto \exp(-i2\pi\nu t)\exp(-i2\pi\alpha x_1)\exp(i\varphi)\sum_j a_j'\sum_k\left[\int_0^t a_j(t')b_k(t')dt'\right]h(x_1-kD',y). \quad (16)$$

For a hologram thickness $n_0L = 1$ cm, placed 1 cm off-axis and 10 cm from input and output neural planes we can separate the etalons by $D \approx 100$ μm, (at the fourth zero of the sinc), allowing 100 to be packed per centimeter. However, for a 90° diffraction angle, as illustrated in Fig. 6, we can bring the etalons to within $D \approx 10$ μm of each other allowing a linear packing density of 1000 etalons/cm. When 2-D arrays of neurons are to be interconnected additional constraints must be imposed on their topology to achieve the appropriate selectivity of the diffracted orders, and only a spare selection of a 2-D array of etalons may be utilized, containing between 10^4 and 10^5 etalons/in.^{2}.[23] An appropriate topology for the utilized etalons can be derived by considering the interconnection to be space variant in the interaction plane and space invariant in the orthogonal dimension. In the case of a Fresnel volume hologram there is some space-variant widening of the impulse response in the direction orthogonal to the interaction plane, which is not present with Fourier holograms.

The resulting field incident on the spatially multiplexed output neurons is found to be proportional to the desired matrix vector product of the input activity pattern a_i' with the time-integrated outer product of the sequence of forward and backward waves. The intensity at the output neural plane is given by the modulus squared of the field, and this intensity will be detected by the neurons and used for subsequent nonlinear processing.

This is a fully self-aligning dynamic volume holographic global interconnection scheme which works reciprocally for forward and backward waves as required by the backpropagation algorithm. This interconnection technique requires no lenses, because the Fresnel hologram accomplishes the imaging operation.

V. Experimental Investigation of Fresnel Volume Hologram Interconnections

Volume Fresnel holograms were recorded in photorefractive crystals as the interference pattern between expanding and collapsing spherical waves to test their capabilities as lensless interconnection elements. When this volume hologram was reilluminated by one or more of an array of expanding reference beams, the collapsing spherical object waves were reproduced, which focused to an array of small spots at the output plane. First, a single input point was interconnected to a single output point using an expanding wave interfered with a collapsing wave in the volume of a $LiNbO_3$ crystal, and the interference pattern was time integrated for several minutes to build up a reasonably high diffraction efficiency grating. A magnified image of the diffracted focal spot which was produced at the output plane when the crystal was reilluminated by the expanding reference beam is shown in Fig. 8. A good focused spot is produced, but when the film is overexposed a large amount of sidelobe structure becomes apparent. A large amount of fanning of the diffracted light is produced in the plane of the crystal c axis due to the recording of additional gratings in the crystal between the reference beam and scattered object beam. A vertical line appears at the output plane which is due to the Bragg matched diffraction of the reference beam by the gratings formed between the scattered reference with the object beam, and this line is actually a small part of a large circle of confusion which passes through the reference source and the object virtual source. These fanning components built up over a longer time scale than the desired focusing diffracted light and were not visible with short holographic exposures. The weak additional spot is a multiple reflection artifact. A measurement of the Bragg positional sensitivity in and out of the principal interaction plane is shown in Fig. 9. In the plane a good approximation to a sinc function with 21-μm width is obtained, which is near the expected width for this experimental geometry.

The measurement was obtained by translating the Fresnel volume hologram and measuring the resulting diffraction efficiency. However, when the hologram was rotated in the plane, and any residual translation was compensated, a diffraction efficiency was measured that was essentially independent of angle, as expected for these angularly diverse volume holograms. Out of the interaction plane the Fresnel hologram diffraction efficiency was quite insensitive to the hologram position. However, the position of the diffracted focused spot translated across the detector array as the hologram was moved, indicating that the holographic interconnection was space invariant in this dimension. A Fresnel hologram that is thick in relation to the separation between planes produces a vertically widening impulse response as the out-of-plane offset is increased due to different offset magnifications at different hologram depths. This feature needs to be considered when selecting a 2-D neuron array topology for use with the Fresnel hologram interconnection scheme.

An optical neural network interconnection pattern requires many point sources to be imaged to many other virtual sources, and the Fresnel hologram was tested in this application by using lenslet arrays for the optical sources. A line array of real sources produced by a 1-D lenslet array was interconnected to a 2-D array of virtual sources that was produced by imaging the focal plane of a lenslet array through and beyond the volume hologram. In this $N \rightarrow N^2$ interconnection experiment approximately fifty sources were interconnected with a 50×50 array of output focal spots, thereby implementing more than 10^5 holographic interconnection lenses. A small portion of the diffracted output plane produced by this hologram when it was illuminated by the object wave is shown in Fig. 10. This looks almost identical to the image of the lenslet array produced by the object beam, and no fanning artifacts like those shown in Fig. 8 are visible. This is because the diffraction efficiency of each interconnection hologram is extremely small in this case, and the weak fanning artifacts produced by different sources do not add up constructively.

Adaptive holographic interconnection networks

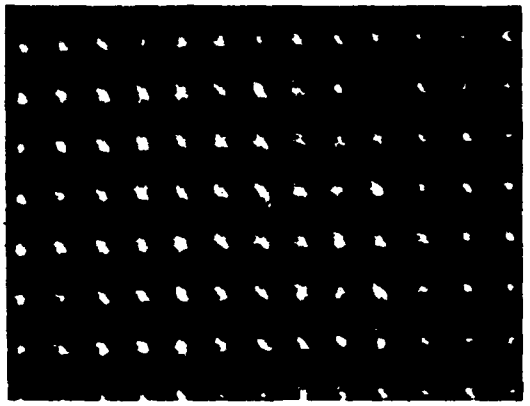


Fig. 10. Diffracted output produced by an $N \rightarrow N^2$ lensless holographic interconnecton.

must be able to represent bipolar weights and to decrease interconnection strengths, so selective erasure was examined as one possible technique that can be used for both these purposes. To show that Fresnel gratings could be selectively erased, $Bi_{12}SiO_{20}$ was used as the holographic recording medium so that photovoltaic effects could be eliminated, and faster response times could be obtained. A piezoelectric mirror was used to phase modulate an object beam with a π phase shift increment. Interconnection gratings were built up with one phase, then the object phase was shifted, and the diffraction efficiency of the reference beam into the object focal spot was measured as the hologram was erased and rewritten with a π phase shift. Because different wavelength probe beams cannot be used to measure the diffraction efficiency of Fresnel holograms, the object beam and diffracted beam were alternatively chopped in a nonoverlapping fashion to measure the diffraction efficiency seen by the reference beam as a function of time. An example of this type of selective erasure process is shown in Fig. 11(b), and it is to be compared with the incoherent erasure that was obtained by blocking the reference beam as shown in Fig. 11(a). The selective erasure was much faster than the incoherent erasure process, or the writing process after the previous grating was erased, because the incoherent erasure and phase shifted writing are cooperating processes during selective erasure, while they are competing processes when writing the hologram. The phase could be repetitively shifted by π as shown in Fig. 11(c), and a succession of out-of-phase gratings can be written and erased. Other gratings within the crystal were not erased any faster with this phase shifted reference approach than they were normally by incoherent erasure, which demonstrates that selective erasure of the Fresnel hologram is occurring throughout the volume of the crystal.

Polarization switching diffraction can be demonstrated in $Bi_{12}SiO_{20}$ by writing a grating in the 110 direction.[27] The propagating eigenmodes are circular without an applied field, because of optical activity, and the right mode can be coupled to the left mode

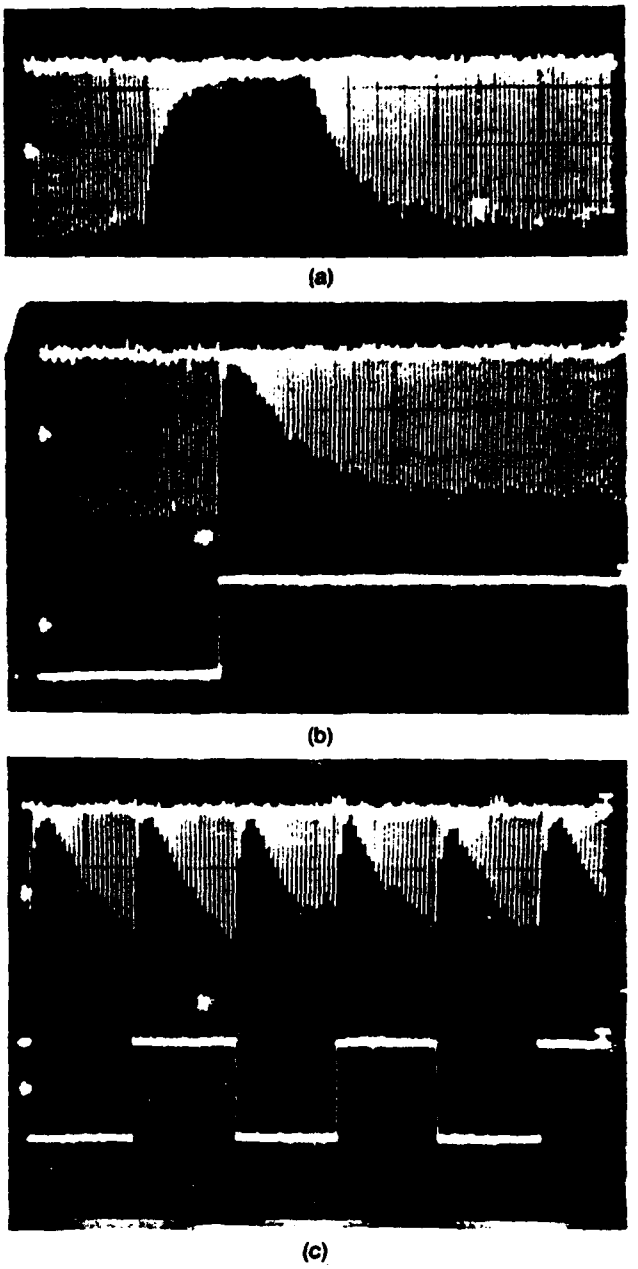


Fig. 11. Erasure processes in $Bi_{12}SiO_{20}$: (a) incoherent erasure process; (b) selective erasure process using a π phase shifted reference and the phase shift signal; (c) repetitive π phase shift writing and erasure (1 s/div).

through the off diagonal tensor components of a photorefractively induced pertubation grating. Right circularly polarized expanding spherical reference waves were interfered with right circularly polarized focusing object beams to record polarization switching Fresnel volume holograms in a properly rotated $Bi_{12}SiO_{20}$ crystal. The diffracted field focused to the object beam focal spot, and the polarization state was analyzed with a properly rotated quarterwave plate and polarizer, and it was found to be very nearly orthogonal to the

polarization of the input object beam. The orthogonality of these modes can be improved by careful alignment of the crystal axes with respect to the principal interaction plane, but perfect orthogonality is probably impossible with a Fresnel hologram because of the angular diversity of the birefringent gratings. When 99% of the object beam can be filtered with this polarization filtering scheme, a 5% diffraction efficiency hologram results in a 20% feedthrough of the undiffracted light measured with respect to the diffracted polarization switched light. This suppression ratio of the undiffracted beam must be improved if the polarization multiplexed architecture is to be used for a backpropagation network, so that the undiffracted phase conjugate reference will not corrupt the diffracted backpropagating error.

VI. System Requirements

A complete two-layer system, illustrated in Fig. 12, requires a high speed method of entering data for pattern transformation processing and another means of introducing the backward-propagating error signals for the learning phase. Probably the best approach to high speed data entry at the back end of the system is a sparse parallel laser diode array or a fiber-optic input array, demagnified onto the first layer nonlinear etalon array which is operated in the bistable regime. In this manner the subthreshold coherent bias beams transmitted by each addressed device can be modulated by the data signals, thereby using the input nonlinear Fabry-Perot etalon array as a high speed incoherent to coherent converter with memory. At the final layer of the system error signals need to be computed and injected back into the system with the appropriate polarization or wavelength and the phase shift or timing needed to represent the sign of the error. The system can be designed with either optical or electronic error detection and generation circuitry at the output to introduce the backpropagating error. Optical subtraction techniques can be considered for an optical approach to teaching the system. Image subtraction using a phase conjugated Michelson interferometer[28] appears to be a promising approach for this application since it produces subtracted fields with the appropriate phase shift to represent the sign of the error, without the accurate phase adjustments required by other interferometric approaches to image subtraction. Alternatively, since the computational load required at the output is relatively minor, optical detectors can be combined with electronic subtraction from the target vector to generate the bipolar error vector, which can be applied to a spatial light modulator at the output to introduce the backpropagating error. When the number of outputs of the pattern transformation procedure is <1000, they can be arrayed in a linear format which allows the utilization of high speed linear detector arrays for output, and the utilization of linear spatial light modulators, to introduce the backward-propagating error signals.

The fan-out capability of each layer is determined by the gain of the nonlinear devices, the holographic diffraction efficiency, and the polarization component throughput, and it will dictate an information collapsing network architecture. For example, if the product of optical efficiencies is only 3%, a network with 30,000-bit input pattern vectors might be processed by 1000 hidden units that communicate with thirty output devices, which simplifies the error generation process at the output. The ability of the system to process large amounts of data in parallel at a very high speed is limited by the electronic addressing of the input array, and the output photodetector array readout time, and not by the intervening optical system, because of the extremely fast response achievable with nonlinear etalons and the almost instantaneous optical interconnection delay. The optical power requirements of the system are primarily dictated by the first layer of nonlinear etalons, since there are many more in this layer than in the succeeding layers for a collapsing network. The first-layer etalons are not bidirectional and can be optimized to have a low switching energy. Bistable nonlinear etalons have been operated with a 3-pJ switching energy at a rate of ~100 MHz,[29] which leads to a power requirement of 0.4-mW/etalon or 12 W for 30,000 input etalons. Only a portion of this power is dissipated within the nonlinear etalons, and a heat dissipation requirement of only a few watts per cm^2 should be achievable with forced air or liquid cooling techniques. Most of this power is supplied by a high power coherent pump beam that is used to bias each bistable device just below the bistable loop, and ~10%, or 40 μW, is required per laser or fiber-optic input to

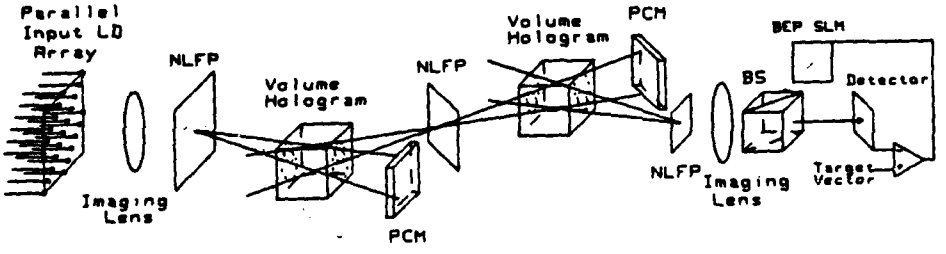


Fig. 12. Complete system for two-layer backpropagation optical learning including massively parallel input laser array and electronic error detection at the output. LD = laser diode (or fiber optics), NLFP = nonlinear Fabry-Perot, PCM = phase conjugate mirror, BEP SLM = spatial light modulator for backward-propagating error.

the network. The backpropagating neurons will require more optical power because the dual functions of the bidirectional cavities conflict with the requirements for a low power device. Since there are not as many hidden and output devices as input devices the systems power requirements are primarily dictated by the size of the input array.

The forward-propagating signal can be a narrow pulse since the response of GaAs nonlinear Fabry-Perot etalon is determined by the peak power incident. In this case the backward-propagating error signal can be either pulsed or cw. In the pulsed mode the PCM would need to have practically instantaneous response, such as a nonlinear optical semiconductor might provide, and the forward- and backward-propagating pulses could be time jittered so they do not overlap in the volume hologram, but the phase conjugate reference and the backward-propagating error pulse would overlap within the crystal, thereby exposing a hologram. Alternatively, the backward-propagating error signal could be a low power cw beam that would not have a high enough peak power to nonlinearly modify the index within the Fabry-Perot etalons, and the forward-propagating pulse could be turned into a quasi cw phase conjugated reference by using a photorefractive crystal-based PCM which has a slow integrated response. The Fabry-Perot etalons would need to have a slow relaxation time of the nonlinearly shifted index (this requires long carrier life times and should lead to lower etalon switching energies); thus the probe beam would have the appropriate response for most of the interval between pulses of the forward beams. In this case the holographic exposure would be due to the time integral of the cw waves in the volume hologram, and the orthogonally polarized and pulsed forward-propagating beam would not contribute significantly to the hologram exposure.

VII. Conclusion

The 3-D storage capacity of volume holograms allows the construction of huge globally interconnected multilayer optical networks which are well beyond the projected capabilities of alternative technologies. The optical system seems well matched to the bidirectional requirements of a backpropagation learning system because of the intrinsically reciprocal nature of optical interconnections. Error driven learning operations, such as backpropagation, should be able to compensate for many of the technological flaws inherent to an optical implementation by adaptively sensing the misbehavior of the system and driving it in the appropriate direction necessary to overcome its imperfections. The nonideal optical implementation of a backpropagation network may actually have improved performance over that of an idealized digital simulation because noise will always be present in the system, helping it to avoid shallow local minima, and pushing the interconnection matrix away from solution boundaries. Imperfections of the holographic interconnection will help the system perform symmetry breaking, which the idealized model cannot perform spontaneously. The simultaneous self-aligning and learning of the optical system make this approach to multilayer optical neural processing experimentally feasible and allow the implementation of complicated systems that could not be completely specified *a priori* but can be learned and modified as the desired processing operation slowly changes. The slow learning of the holographic crystals combined with the extremely high speed processing of the nonlinear etalons gives this system an enormous throughput potential and the capability for solving complicated cognitive problems.

The authors would like to acknowledge the numerous contributions to this work made by David Brady as well as useful discussions with Jeff Yu and Hyatt Gibbs.

The work reported here was partially supported by DARPA, the Army Research Office, and the Air Force Office of Scientific Research.

References

1. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing, Vol. 1*, D. E. Rumelhart and J. L. McClelland, Eds. (MIT Press, Cambridge, MA, 1986), Chap. 8.
2. D. B. Parker, "Learning Logic," Invention Report S81-64, File 1, Office of Technology Licensing, Stanford U. (Oct. 1982).
3. J. J. Hopfield, "Neurons with Graded Response have Collective Computational Properties like those of Two-State Neurons," Proc. Natl. Acad. Sci. USA 81, 3088 (1984).
4. S. Grossberg, *Studies of Mind and Brain* (Reidel, Boston, 1982).
5. T. Kohonen, *Self-Organization and Associative Memory* (Springer-Verlag, Berlin, 1984).
6. Y. S. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," Sci. Am. 256, 88 (1987).
7. D. Psaltis and N. H. Farhat, "Optical Information Processing Based on an Associative Memory Model of Neural Nets with Thresholding and Feedback," Opt. Lett 10, 98 (1985).
8. D. Z. Anderson, "Coherent Optical Eigenstate Memory," Opt. Lett. 11, 56 (1986).
9. B. H. Soffer, G. J. Dunning, Y. Owechko, and E. Marom, "Associative Holographic Memory with Feedback Using Phase-Conjugate Mirrors," Opt. Lett. 11, 118 (1986).
10. A. Yariv and S. Kwong, "Associative Memories Based on Message-Bearing Optical Modes in Phase Conjugate Resonators," Opt. Lett. 11, 186 (1986).
11. T. Jannson *et al.*, "The Interconnectability of Neuro-Optic Processors," Proc. Soc. Photo-Opt. Instrum. Eng. 698, 157 (1986).
12. A. D. Fisher et al., "Implementation of Adaptive Associative Optical Computing Elements," Proc. Soc. Photo-Opt. Instrum. Eng. 525, 196 (1986).
13. M. Cohen, "Design of a New Medium for Volume Holographic Information Processing," Appl. Opt. 25, 2288 (1986).
14. N. Farhat, "Architectures for Opto-Electronic Analogs of Self-Organizing Neural Networks," in *Technical Digest of Topical Meeting on Optical Computing* (Optical Society of America, Washington, DC, 1987), p. 125.
15. K. Wagner and D. Psaltis, "Multilayer Optical Learning Networks," Proc. Soc. Photo-Opt. Instrum. Eng. 752, 16 (1987).
16. K. Wagner and D. Psaltis, "Multilayer Optical Learning Networks," in *Technical Digest, Topical Meeting on Optical Computing* (Optical Society of America, Washington, DC, 1987), p. 133.
17. D. Psaltis and C. Park, "Nonlinear Discriminant Functions and Associative Memories in *Proceedings, Conference on Neural*

Networks for Computing, Snowbird, UT, APS Conf. Proc. 151 (1986).
18. T. J. Sejnowski and C. R. Rosenberg, "NETtalk: a Parallel-Network that Learns to Read Aloud," John Hopkins U., JHU/EECS-86/01 (1986).
19. B. Widrow and M. E. Hoff, "Adaptive Switching Circuits," IRE Wescon Conv. Rec. 4, 96 (1960).
20. H. M. Gibbs, *Optical Bistability: Controlling Light with Light* (Academic, New York, 1985).
21. H. M. Gibbs, *et al.*, "Optical Modulation by Optical Tuning of a Cavity," Appl. Phys. Lett. 34, 511 (1979).
22. A. W. Lohmann, "Polarization and Optical Logic," Appl. Opt. 25, 1594 (1986).
23. D. Psaltis *et al.*, "Optical Neural Nets Implemented with Volume Holograms," in *Technical Digest Topical Meeting on Optical Computing* (Optical Society of America, Washington, DC, 1987).
24. J. P. Huignard *et al.*, "Coherent Selective Erasure of Superimposed Volume Holograms in $LiNbO_3$," Appl. Phys. Lett. 26, 256 (1975).
25. D. L. Stabler *et al.*, "Multiplier Storage and Erasure of Fixed Holograms in Fe-Doped $LiNbO_3$," Appl. Phys. Lett. 26, 182 (1975).
26. J. D. Cresser and P. Meystre, "The Role of Phases in the Trasient Dynamics of Nonlinear Interferometers," in *Optical Bistability*, C. M. Bowden Ed. (Plenum, New York, 1980).
27. A. Merrakchi, R. V. Johnson, and A. Tanguay, Jr., "Polarization Properties of Photorefractive Diffraction in Electrooptic and Optically Active Sillenite Crystals (Bragg Regime)," J. Opt. Soc. Am. B. 3, 321 (1986).
28. A. E. Chiou and P. Yeh, "Parallel Image Subtraction Using a Phase Conjugate Michelson Interferometer," Opt. Lett. 11, 306 (1986).
29. J. L. Jewell *et al.*, "3pJ 82MHz Optical Logic Gates in a Room Temperature GaAs-AlGaAs Multiple Quantum Well Etalon," Appl. Phys. Lett. 46, 918 (1985).

Reprinted from Applied Optics, Vol. 27 page 1752, May 1, 1988

Adaptive optical networks using photorefractive crystals

Demetri Psaltis, David Brady, and Kelvin Wagner

The capabilities of photorefractive crystals as media for holographic interconnections in neural networks are examined. Limitations on the density of interconnections and the number of holographic associations which can be stored in photorefractive crystals are derived. Optical architectures for implementing various neural schemes are described. Experimental results are presented for one of these architectures.

I. Introduction

Learning is the most distinctive feature of a neural computer and in many respects it is this aspect that gives neural computation an advantage over alternative computational strategies. A neural computer is trained to produce the appropriate response to a class of inputs by being presented with a sufficient number of examples during the learning phase. The presentation of these examples causes the strength of the connections between neurons that comprise the network to be modified according to the specifics of the learning algorithm. A successful learning procedure will result in a trained network that responds correctly when it is presented with the examples it has seen previously and also other inputs that are in some sense similar to the known patterns. When we consider a physical realization of a neural network model, we have two options in incorporating learning capability. The first is to build a network with fixed but initially programmable connections. An auxiliary, conventional computer can then be used to learn the correct values of the connection strengths and once learning has been completed the network can be programmed by the computer. While this approach may be reasonable for some applications, a system with continuously modifiable connections presents a much more powerful alternative.

In this paper we consider the optical implementation of learning networks using volume holographic interconnections in photorefractive crystals. The use of volume holograms permits the storage of a very large number of interconnections per unit volume,[1-4] whereas the use of photorefractive crystals permits the dynamic modification of these connections, thus allowing the implementation of learning algorithms.[5-9] We first briefly review the major types of learning algorithms that are being used in neural network models. We then estimate the maximum number of holographic gratings that can simultaneously exist in a photorefractive crystal. Since in an optical implementation each grating corresponds to a separate interconnection between two neurons, this estimate gives us the density of connections that are achievable with volume holograms. The next topic that we address is how the modulation depth of each grating (or equivalently the strength of each connection) can be controlled through the implementation of learning algorithms. Two related issues are investigated: the optical architectures which implement different learning algorithms and the reconciliation of physical mechanisms that are involved in the recording of holograms in photorefractive crystals with the dynamics of the learning procedures in neural networks.

II. Learning Algorithms

For the purposes of this discussion it is convenient to separate the wide range of learning algorithms that have been discussed in the literature into three categories: prescribed learning, error driven learning, and self-organization. We will draw the distinction among these with the aid of Fig. 1, where a general network is drawn with the vector $\mathbf{x}(k)$ as its input and $\mathbf{y}(k)$ the output at the kth iteration (or time interval). The vector $\mathbf{z}(k)$ is used to represent the activity of the internal units and $w_{ij}(k)$ is the connection strength between the ith and the jth units. Let $\mathbf{x}^{(m)}$, $m = 1 \ldots M$, be a set of specified input vectors and let $\mathbf{y}^{(m)}$ be the responses which the network must produce for each of these input vectors.

The authors are with California Institute of Technology, Pasadena, California 91125.
Received 24 July 1987.
0003-6935/88/091752-08$02.00/0.

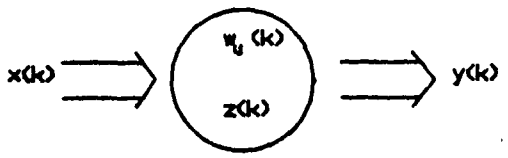


Fig. 1. General neural network architecture.

A prescribed learning algorithm calculates the strength of each weight simply as a function of the vectors $\mathbf{x}^{(m)}$ and $\mathbf{y}^{(m)}$:

$$w_{ij} = f_{ij}[\mathbf{x}^{(m)},\mathbf{y}^{(m)}] \qquad m = 1 \ldots M. \tag{1}$$

This type of procedure is relatively simple (easy learning). It is perhaps the most sensible approach in a single layer network. The widely used outer product algorithm[10,11] is an example of this type of learning algorithm, as are some schemes which utilize the pseudoinverse.[10–13] Despite its simplicity, prescribed learning is limited in several important respects. First, while prescribed learning is well understood for single layer systems, the existing algorithms for two layers are largely localized representations; each input $\mathbf{x}^{(m)}$ activates a single internal neuron.[14–16] Moreover, the entire learning procedure usually has to be completed *a priori*. This last limitation is not encountered in the simplest form of prescribed learning, the outer product rule:

$$w_{ij} = \sum_{m=1}^{M} x_i^{(m)} y_i^{(m)}. \tag{2}$$

In this case new memories may be programmed by simply adding the outer products of new samples to the weight matrix. Note that once the interconnection matrix has been determined by a prescribed learning algorithm, it may be expressed in the form of a sum of at most N outer products, where N is the total number of neurons in each layer. Since volume holograms record interconnection matrices represented by sums of outer products in a very natural way, matrices which can be expressed in this form are particularly simple to implement in optics.[17–20]

Error driven learning is distinguished by the fact that the output of the system, $\mathbf{y}(k)$, is monitored and compared to the desired response $\mathbf{y}^{(m)}$. An incremental change is then made to the interconnection weights to reduce the error:

$$\Delta w_{ij}(k) = f_{ij}[\mathbf{x}^{(m)},w_{rs}(k),\mathbf{y}^{(m)}]. \tag{3}$$

The change Δw_{ij} is calculated from the vectors $\mathbf{x}^{(m)}$ and $\mathbf{y}^{(m)}$ and the current setting of the weight matrix $w_{rs}(k)$ (from which the state of the entire network can be calculated). The perceptron[21] and adaline[22] algorithms are examples of error driven learning for single layer networks. Interest in such learning algorithms has been renewed recently by the development of procedures suitable for multilayered networks.[23–25] Error driven algorithms (hard learning) are more difficult to implement than prescribed learning since they require a large number of iterations before errors can be reduced to sufficiently low levels. In multilayered systems, however, this type of learning can provide a effective mechanism for matching the available re sources (connections and neurons) to the requirement of the problem. In optical realizations error drive algorithms are more difficult to implement than pre scribed approaches due to the need for dynamicall modifiable interconnections and the incorporation o an optical system that monitors the performance an causes the necessary changes in the weights.[26] Whil this problem could be avoided by performing learnin off line in computer simulations and recording th optimized interconnection matrix as in prescribe learning, this approach has the disadvantage that onc again the matrix is fixed *a priori*, thus preventing th network from being adaptive. In subsequent section we will consider a relatively simple form of Eq. (3) i which $\Delta w_{ij}(k)$ depends only on locally available infor mation, i.e., z_i in one layer and z_j in an adjacent laye

$$\Delta w_{ij}(k) = f_{ij}[z_i[w_{rs}(k),\mathbf{y}^{(m)},\mathbf{x}^{(m)}],z_j[w_{rs}(k),\mathbf{y}^{(m)},\mathbf{x}^{(m)}]]. \tag{4}$$

The perceptron and the backward error propagatio algorithms both fall in this subcategory if we allow th neuronal activity z_i to include error signals, i.e., if eac neuron has distinct signal and error outputs which a separated temporally or spatially. An example of suc a neuron implemented in optics is given below in co junction with an optical back error propagation sy tem.

In the case of self-organizing learning algorithms require not that the specified inputs produce a partic lar response but rather that they satisfy a gener restriction, often imposed by the structure of the ne work itself. Since there is no *a priori* expected r sponse, the learning rule for self-organizing systems simply

$$\Delta w_{ij}(k) = f_{ij}[\mathbf{x}^{(m)},w_{rs}(k)].$$

This type of learning procedure can be useful, f example, at intermediate levels of a network where t purpose is not to elicit an external response but rath to generate appropriate internal representations of t information that is presented as input to the netwo There is a broad range of self-organizing algorithn the simplest of which is probably lateral inhibition enforce grandmother cell representations.[10,27] T objective of the learning procedure is to have ea distinct pattern in an input set of neurons activat single neuron in a second set. In the architect shown in Fig. 2 this is accomplished via inhibit connections between the neurons in the second s Once a particular neuron in the second layer is parti ly turned on for a specific pattern it prevents t connections to the other neurons in the second set fr assuming values that will result in activity at m than one neuron. The details of the dynamics of s procedures can be quite complex (e.g., see Ref. 28), can corresponding optical implementations. An vantageous feature of optics in connection with s organization is that global training signals, such fixed lateral inhibition between all the neurons i given layer, can easily be broadcast with optical bea

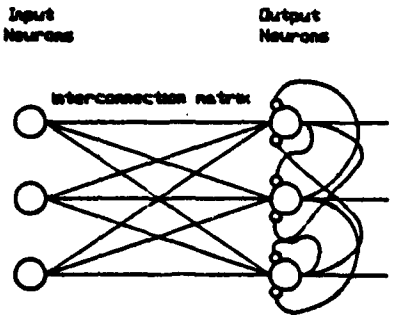


Fig. 2. Two-layer network with lateral inhibition. Connections ending with an open circle are inhibitory.

III. Interconnection Capabilities of Volume Holograms

The basic architecture for optical implementation of a neural computer is shown in Fig. 3. The figure presents a single stage of what may be a multilayered system. The nonlinear processing elements (i.e., the neurons) are arranged in planes. We have included a training plane for reasons which will become clear below. Neurons in one plane are interconnected with the neurons in the same or other planes via the third dimension. The strength of the interconnections is determined by the information which is holographically stored in light sensitive media placed in the space separating the neural planes. Volume, rather than thin, holograms are specified in Fig. 3 due to the much greater storage capacity of volume holograms and the availability of excellent real-time volume media. Photorefractive crystals are particularly attractive as holographic media in this application because it is possible to record information in these crystals in real time at very high density without degrading the photorefractive sensitivity. In this section we discuss the factors that determine the maximum number of connections that can be specified by a photorefractive crystal with a given set of physical characteristics. There are three distinct factors that need to be considered: geometric limitations arising from the basic principles of volume holography, limitations rising from the physics of photorefractive recording, and limitations due to the learning algorithms.

The Fourier lenses in Fig. 3 transform the spatial position of each neuron into a spatial frequency associated with light emitted by or incident on that neuron. An interconnection between the ith neuron in the input plane and the jth neuron in the output plane is formed by interfering light emitted by the input neuron with light emitted by the jth neuron in the training plane. The image of the jth training neuron lies at the position of the jth neuron in the output plane. The interference of the training signal and the input creates a grating in the recording medium of the form

$$\Delta\chi_{ij} = A_i A_j^* \exp(j\mathbf{K}_{ij}\cdot\mathbf{r}), \tag{6}$$

where A_i and A_j are the amplitudes of the fields emitted by the ith and jth neurons, respectively. $\mathbf{K}_{ij}$ is equal to $\mathbf{k}_i - \mathbf{k}_j$ where $\mathbf{k}_i$ and $\mathbf{k}_j$ are the spatial frequencies at which the corresponding amplitudes propagate

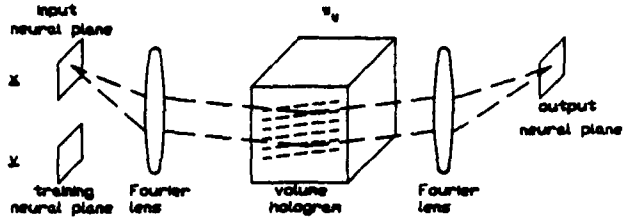


Fig. 3. Optical neural computer architecture.

in the volume medium. This grating diffracts an input beam at spatial frequency $\mathbf{k}_\alpha$ into an output beam at spatial frequency $\mathbf{k}_\beta$ if these two beams satisfy the Bragg constraint that

$$\mathbf{k}_\alpha - \mathbf{k}_\beta = \mathbf{K}_{ij}. \tag{7}$$

This constraint is obviously satisfied if $\mathbf{k}_\alpha = \mathbf{k}_i$ and $\mathbf{k}_\beta = \mathbf{k}_j$. In general this solution is not unique. However, Psaltis *et al.*[2,3] have shown that by placing the neurons on the input and output planes on appropriate fractal grids of dimension 3/2 it is possible to insure that only the ith input neuron and the jth output neuron may be coupled by a grating with wave vector $\mathbf{K}_{ij}$. In this case, recording a hologram between light from the ith input neuron and the jth training neuron increases the connection strength between the ith input and the jth output without directly affecting the connections between other neurons. If instead of one neuron, patterns of neurons are active on the fractal grids of the input and training planes, the hologram recorded in the volume, i.e., Eq. (6) summed over all active pairs of neurons, is the outer product of the pattern on the input plane and the pattern on the training plane. Exposing the hologram with a series of M pattern yields the sum of outer products described by Eq. (2). Note that the architecture shown in Fig. 3 is similar to a joint Fourier transform correlator. The use of volume, rather than thin, holograms and fractal grids destroys the shift invariance of the correlator, making this architecture a totally shift-variant arbitrarily interconnectable system.

A basic geometrical limitation on the density of interconnections achievable through volume holograms is due to the finite volume V of any real crystal. The refractive index $n(\mathbf{r})$ of such a crystal under periodic boundary conditions may be represented in the form

$$n(\mathbf{r}) = \sum_{\nu}^{S} n_\nu \exp(j\mathbf{k}_\nu\cdot\mathbf{r}), \tag{8}$$

$$\mathbf{k}_\nu = \left[\nu_x\left(\frac{2\pi}{L_x}\right)\hat{x} + \nu_y\left(\frac{2\pi}{L_y}\right)\hat{y} + \nu_z\left(\frac{2\pi}{L_z}\right)\hat{z}\right]$$

$$\nu_i = 0,\pm1,\pm2\ldots, \tag{9}$$

where n_ν is the amplitude of the Fourier component at spatial frequency $\mathbf{k}_\nu$ and L_i is the length of the crystal in the $\hat{i}$ direction. Since the maximum spatial frequency which may be Bragg matched to diffract light at wavelength λ is $2k_0$, where $k_0 = 2\pi/\lambda$, the sum in Eq. (8) is finite in holographic applications. The number of spatial frequencies in the sum is $S \approx V/\lambda^3$. Psaltis *et*

al.[2,3] demonstrated that S is sufficient to fully and independently interconnect neural planes which are limited to fractal dimension 3/2. Thus in this previous work the issue of these geometric limitations was fully resolved in the condition that processing nodes in the input and output planes must be appropriately arranged on fractal grids. Other geometric limitations arise due to finite numerical apertures and the physics of holographic recording mechanisms. These factors may be shown to contribute a scaling factor to S which is independent of V and λ. For $V = 1$ cm^3 and $\lambda = 1$ μm, V/λ^3 is equal to 10^{12}. In interconnecting neurons arranged on fractal planes, even though the recording geometry typically allows access to only 1% of grating wave vector space, we still may achieve 10^{10} interconnections per cm^3.

We now address the question of whether this large number of gratings can be supported in a photorefractive crystal, i.e., do photorefractive crystals have the capability of simultaneously storing 10^{10} gratings each with sufficient diffraction efficiency? In this paper we answer this question based on simple arguments in the context of a neural architecture. The conclusions we reach are the same as those we arrive at through a more thorough examination of the problem. Photorefractive holograms are produced in electrooptic crystal via the modulation of the index of refraction by the space charge field created by an optically driven inhomogeneous charge distribution. A neural network architecture implemented in volume holograms performs a transformation of the form

$$\begin{aligned} E_{i\,\mathrm{in}} \exp(j\mathbf{k}_i \cdot \mathbf{r}) \exp(j\phi_i) + \text{c.c.} = \sum_j \eta_{ij} \exp(j\psi_{ij}) \\ \times \exp(j\mathbf{K}_{ij} \cdot \mathbf{r}) \\ \times E_{j\,\mathrm{out}} \exp(j\mathbf{k}_j \cdot \mathbf{r}) \\ \times \exp(j\phi_j) + \text{c.c.} \end{aligned} \tag{10}$$

between the field amplitude, $E_{j\,\mathrm{out}} \exp(j\mathbf{k}_j \cdot \mathbf{r})$, of the jth neuron and the field amplitude, $E_{i\,\mathrm{in}} \exp(j\mathbf{k}_j \cdot \mathbf{r})$, incident on the input of the ith neuron. c.c. denotes the complex conjugate of the preceding term. ϕ_j and ϕ_i are the phases of the field amplitudes corresponding to the ith and jth neurons. ψ_{ij} is the phase of the grating which connects the ith and jth neurons. The field amplitude diffraction efficiencies η_{ij} are proportional to the component of the space charge density in the crystal at spatial frequency $\mathbf{K}_{ij} = \mathbf{k}_i - \mathbf{k}_j$.[29] The total space charge density due to N stored gratings is constrained at every point in the crystal to be less than the acceptor trap density. This implies that

$$\mathcal{R}\left\{\sum_i \sum_j \eta_{ij} \exp(j\psi_{ij}) \exp(j\mathbf{K}_{ij} \cdot \mathbf{r})\right\} \leq \eta_0, \tag{11}$$

where η_0 is the maximum diffraction efficiency for the field amplitude when only one grating is recorded. If ψ_{ij} is an independent uniformly distributed random variable on $(-\pi,\pi)$, with high probability the right-hand side of Eq. (11) will not exceed a few times its standard deviation, $\sqrt{N/2}\eta_1$, where η_1 is the rms value of η_{ij}. This fact allows us to find a simple limit for η_1 given by

$$\eta_1 \approx \frac{\eta_0}{\sqrt{\frac{N}{2}}}. \tag{12}$$

Note that, although we have assumed that the sums in Eq. (11) are over a set of incoherent sinusoids, this does not imply that the sum in Eq. (10) is incoherent. To illustrate this point imagine that $\psi_{ij} = \phi_i - \phi_j$. In this case the terms in Eq. (10) add coherently. However if ϕ_i and ϕ_j are independent random variables the sums in Eq. (11) still add incoherently. Thus a random phase term in the transmittance at each neuron causes the charge densities stored in the crystal to add incoherently but does not necessarily destroy the coherence of the optical system.

The holographic transformation described above can be used to implement neural architectures which map an activity pattern described by the outputs $\{x_j\}$ of the neurons on one neural plane to the outputs $\{y_i\}$ of the next neural plane. In a coherent optical system x_j is represented by $E_{j\,\mathrm{out}} \exp(j\phi_j)$ and w_{ij} is represented by $\eta_{ij} \exp(j\psi_{ij})$. Since most simple optical nonlinearities are based on absorption the transformation between $\{x_j\}$ and $\{y_i\}$ typically takes the form

$$y_i = f\left(\left|\sum_j w_{ij} x_j\right|^2\right), \tag{13}$$

where f is a thresholding function implemented in the neural plane. This functional form might be avoided using interferometric detection. In an incoherent optical system x_j is represented by $|E_{j\,\mathrm{out}}|^2$ and w_{ij} is represented by η_{ij}^2. The transformation between $\{x_j\}$ and $\{y_i\}$ takes the form

$$y_i = f\left(\sum_j w_{ij} x_j\right). \tag{14}$$

In either case the function f must provide sufficient gain G to regenerate the signal power of the system after each layer. If we assume that each layer contains $\sqrt{N}$ neurons, the relationship between the power incident on a single neuron, I_{in}, and the power output by a single neuron, I_{out}, for a coherent system with $\psi_{ij} = \phi_i - \phi_j$ is

$$\begin{aligned} I_{\mathrm{in}} &= \kappa \left|\sum_j^{\sqrt{N}} \eta_{ij} \exp(j\psi_{ij}) E_{j\,\mathrm{out}} \exp(j\phi_j)\right|^2 \\ &= N\eta_1^2 I_{\mathrm{out}} = \frac{I_{\mathrm{out}}}{G_{\mathrm{coherent}}}. \end{aligned} \tag{15}$$

From Eq. (12) we find

$$G_{\mathrm{coherent}} = \frac{1}{2\eta_0}. \tag{16}$$

For an incoherent system the corresponding relationship is

$$I_{\mathrm{in}} = \kappa \sum_j^{\sqrt{N}} \eta_{ij}^2 |E_{j\,\mathrm{out}}|^2 = \sqrt{N}\eta_1^2 I_{\mathrm{out}} = \frac{I_{\mathrm{out}}}{G_{\mathrm{incoherent}}}. \tag{17}$$

In this case Eq. (12) yields

$$G_{\text{incoherent}} = \frac{\sqrt{N}}{2\eta_0}. \tag{18}$$

Note that $1/G$ is the total diffraction efficiency of the volume hologram. Since this must be less than 1 we know that $G > 1$. η_0 is determined by the physical properties of the crystal, including the maximum charge density available for grating storage, the thickness of the crystal, and its electrooptic coefficients. For small η_1 we may estimate η_0 as

$$\eta_0 \approx \Delta\epsilon \frac{2\pi}{\lambda} L,$$

where L is the length of the crystal along the optical axis. For $\Delta\epsilon \approx 10^{-5}$, $\lambda \approx 10^{-6}$ m, and $L \approx 10^{-2}$ m, $\eta_0 \approx 0(1)$. This means that in coherent systems relatively little gain [i.e., $G \approx 0(1)$] is needed to recall a large number of sinusoidal gratings stored in a photorefractive crystal. Of course as we attempt to store arbitrarily many gratings other limits arise, but at least over a finite bandwidth of the electrooptic response of the crystal coherent systems should have no difficulty in achieving interconnection densities of the order of those implied by the geometrical constraints. Incoherent systems, on the other hand, are unable to take advantage of holographic phase matching and are thus less efficient.[30] To achieve $N = 10^{10}$, for example, we must supply a gain of $G = 10^5$ in each neural plane. Examples of how G may be obtained optically include various combinations of image intensifiers and spatial light modulators and multiwave mixing in nonlinear materials. For example, an optically addressed spatial light modulator such as the Hughes liquid crystal light valve is sensitive to $\sim 10\,\mu\text{W/cm}^2$. If the read-out beam has an intensity of 1 W/cm^2 we achieve a gain of 10^5.

The choice between coherent and incoherent implementations of optical neural networks offers advantages and disadvantages on both sides. The incoherent system is easier to implement but requires the large gain described above and offers only unipolar activities and interconnection strengths. The coherent implementation offers bipolar activities and interconnections but requires rigid phase stability in the optical system over potentially very long learning cycles. This stability is not difficult to achieve in prescribed learning architectures, but may be more difficult to achieve in adaptive systems. In addition, coherent systems generally square the signal incident on the nonlinearity, unless interferometric detection is used. Interferometric detection is difficult to implement in a complex optical system. Although the incoherent system is straightforward to implement, this simplicity comes at a cost of requiring biasing to compensate for unipolar values and external gain. The coherent system is more elegant in that these additional mechanisms are not necessary, but it is more sensitive to specific design issues. One way of making coherent implementations more robust might be to include adaptive optics, such as phase conjugate devices, to compensate for phase instabilities. Although these devices might also be needed in adaptive incoherent systems to detect the phase of a grating to correctly update the associated interconnection, in the incoherent case it is only necessary to detect the current state of the phase. In the coherent case it is generally necessary to continuously track the phase.

IV. Learning Architectures

We now turn to the question of how we can specify the strength of each interconnection. There is a nice compatibility between simple (multiplicative) Hebbian learning and holography; the strength of the connection between two neurons can be modified by recording a hologram with light from the two neurons. It is not possible, however, to record multiple holograms in a single crystal independently. Thus far we have shown that the space charge in a photorefractive crystal may be arranged to achieve a very large number of independent interconnections. The task that remains is to find a means of using optical beams from outside the crystal to correctly arrange the 3-D charge distribution. In particular, we must find means to address the full 3-D bandwidth of the crystal from 2-D neural planes. To successfully implement learning with photorefractive crystals the nonlinear dynamics that govern the multiple exposure of holograms in a photorefractive medium must be reconciled with the nonlinear equations that describe the iterative procedures of learning algorithms. It is extremely difficult to fully characterize analytically the ability of an optical system to simulate a particular learning algorithm. We will have to rely heavily on experiment in the search for the optimum match between nonlinear optics and learning procedures for neural networks. In this section we describe learning architectures which are relatively simple to implement experimentally and which can be used to evaluate the capability of photorefractive crystals to store information in the form of connectivity patterns in a neural computer.

The first learning algorithm we consider is the prescribed sum of outer products of Eq. (2). As we saw in the previous section, a sum of this sort may be implemented as a series of exposures of a volume hologram. In a photorefractive crystal, the exposure of a new hologram partially erases previously recorded holograms. This places an upper limit on the maximum number of holograms that can be recorded and thus the number of associations M that can be stored in the crystal. The limit is found by determining the minimum tolerable diffraction efficiency for each association and solving for the number of exposures that will yield this efficiency. Let A_m be the amplitude of the mth hologram recorded. After a total of M exposures,

$$A_m = A_0\left[1 - \exp\left(\frac{-t_m}{\tau_r}\right)\right]\exp\left(-\sum_{m'=m+1}^{M}\frac{t_{m'}}{\tau_e}\right), \tag{19}$$

where A_0 is the saturation amplitude of a hologram recorded in the photorefractive crystal, t_m is the exposure time for the mth hologram, τ_r and τ_e are, respectively, the characteristic time constants for recording and erasing a hologram in the crystal. We allow for the

case that $\tau_e \neq \tau_r$ in light of limited evidence that this may be the case in some crystals.[31] Ionic conductivity is one mechanism leading to multiple time constants. We can use several different criteria for selecting the exposure schedule t_m. For example, if we require $A_m = A_{m+1}$ for all m we obtain

$$\left[1 - \exp\left(\frac{-t_m}{\tau_r}\right)\right] \exp\left(\frac{-t_{m+1}}{\tau_e}\right) = \left[1 - \exp\left(\frac{-t_{m+1}}{\tau_r}\right)\right]. \quad (20)$$

If $\tau_r = \tau_e$, the solution to Eq. (20) in the boundary condition $t_1 \gg \tau_r$ is

$$t_m = \tau_e \ln\left(\frac{m}{m-1}\right) \qquad m > 1, \quad (21)$$

which yields

$$A_m = A_M = \frac{A_0}{M}. \quad (22)$$

For the case $\tau_r \neq \tau_e$ we define ρ_m such that $t_m = \rho_m \tau_e$. Since, from Eq. (19), $\lim_{M \to \inf} A_1 = 0$, Eq. (20) may be satisfied only if $\lim_{m \to \inf} t_m = 0$. Thus for some $m_0 > 1$, $\rho_{m_0} \ll 1$ and $t_{m_0} \ll \tau_r$. Then, from Eq. (20),

$$t_{m_0+1} \approx \frac{t_{m_0}}{1 + \frac{t_{m_0}}{\tau_e}} = \left(\frac{\rho_{m_0}}{1 + \rho_{m_0}}\right)\tau_e, \quad (23)$$

or

$$\rho_{m_0+1} = \frac{\rho_{m_0}}{1 + \rho_{m_0}}. \quad (24)$$

By induction, for $m > m_0$

$$\rho_m = \frac{1}{(m - m_0) + \frac{1}{\rho_{m_0}}}. \quad (25)$$

As m grows large with m_0 fixed, Eq. (25) can be shown to yield

$$\rho_m \approx \frac{1}{m}, \quad (26)$$

$$t_m = \frac{\tau_e}{m}. \quad (27)$$

The value of m for which the approximation holds increases with the ratio τ_e/τ_r. In the case $\tau_r = \tau_e$, for example, $\tau_e/3t_3 = 0.82$ and $\tau_e/10t_{10} = 0.95$. In any case, for $M \gg m_0$ for some m_0 satisfying the constraints preceding Eq. (23),

$$A_m = A_M = A_0\left[1 - \exp\left(\frac{-\tau_e}{M\tau_r}\right)\right] \quad (28)$$

for all m. Solving for M with $A_m \ll A_0$ we find a limit for M given by

$$M \approx \frac{\tau_e}{\tau_r}\frac{A_0}{A_m}. \quad (29)$$

This result agrees well with what we might expect intuitively. The number of exposures allowed increases in proportion with the ratio τ_e/τ_r (if we erase slowly we can store more holograms) and the ratio of the maximum possible and minimum detectable grating amplitudes.

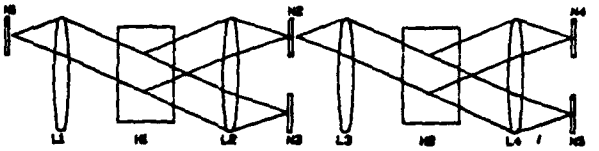


Fig. 4. Optical architecture for backward error propagation learning.

The second architecture we will discuss is capable of implementing the backward error propagation algorithm[23,24] in a multilayered network. The architecture, shown in Fig. 4, is a variation on a system we described previously.[6,8] The system as shown has two layers but an arbitrary number of layers can be implemented as a straightforward extension. An input training pattern is placed at plane N_1. The pattern is then interconnected to the intermediate (hidden) layer N_2 via the volume hologram H_1. A 2-D spatial light modulator placed at N_2 performs a soft thresholding operation on the light incident on it, simulating the action of a 2-D array of neurons, and relays the light to the next stage. Hologram H_2 interconnects N_2 to the output plane N_4 where a spatial light modulator performs the final thresholding and produces a 2-D pattern representing the response of the network to the particular input pattern. This output pattern is compared to the desired output and the appropriate error image is generated (either optically or with the aid of an image detector and rerecording) on the spatial light modulator N_4. The undiffracted beams from N_1 and N_2 are recorded on spatial light modulators at N_3 and N_5, respectively. The signals stored at N_3, N_4, and N_5 are then illuminated from the right so that light propagates back toward the left. The backpropagation algorithm demands a change in the interconnection matrix stored in H_2 given by

$$\Delta w_{ij}^{(2)} = -\alpha \varepsilon_i f'(x_i^{in}) x_j^{out}, \quad (30)$$

where α is a constant, ε_i is the error signal at the ith neuron in N_4, x_i^{in} is the input diffracted onto the ith neuron in N_4 from N_2, $f'(x)$ is the derivative of the thresholding function $f(x)$ which operates on the input to each neuron in the forward pass, and x_j^{out} is the output of the jth neuron in N_2. Each neuron in N_4 is illuminated from the right by the error signal ε_i and the backward transmittance of each neuron is proportional to the derivative of the forward output evaluated at the level of the forward propagating signal. As we have described above, the hologram recorded in H_2 is the outer product of the activity patterns incident from N_4 and N_5. Thus the change made in the holographic interconnections stored in H_2 is proportional to the change described by Eq. (30).

The change in the interconnection matrix stored in H_1 required under the backpropagation algorithm is

$$\Delta w_{lm}^{(1)} = -\sum_i \alpha \varepsilon_i f'(x_i^{in}) w_{il}^{(2)} f'(x_l^{in}) x_m^0, \quad (31)$$

where x_m^0 is the activity on mth input on N_1. The error

signal applied to N_4 produces a diffracted signal at the lth neuron in N_2 which is proportional to

$$-\sum_i \epsilon_i f'(x_i^{in}) w_{il}^{(2)}. \tag{32}$$

We assume that, during the correction cycle for H_1, N_5 is inactive. Once again, if the backward transmittance of the lth neuron is proportional to $f'(x_l^{in})$, the change made to the hologram by the signals propagating back from N_2 and N_3 is proportional to the change prescribed in Eq. (31).

A key element in this architecture is the assumption that the spatial light modulators at N_2 and N_4 may have transmittances which may be switched between a function $f(x)/x$ for the forward propagating signal and $f'(x)$ for the backpropagating signal. In both cases x represents the forward propagating signal. We have previously described how nonlinear etalon switches might be used in this application.[7,8] Electrooptic spatial light modulators might also be used.[8]

We have performed an experiment to show how a single layer of error driven learning might be implemented. This experiment is shown schematically in Fig. 5. In this case, the stored vectors $\mathbf{x}^{(m)}$ correspond to 2-D patterns recorded on a liquid crystal light valve from a video monitor. The output vectors $\mathbf{y}^{(m)}$ correspond to the single bit output of the detector D. An input vector is imaged onto a photorefractive crystal via two separate paths. The strength of the grating between the image of the input along one path and the image along the other path is read out by light propagating along the path of one of the write beams in the orthogonal polarization, i.e., while the write beam incident on the detector is linearly polarized, the other write beam is circularly polarized. The polarizer P blocks the linearly polarized beam and one component of the diffracted circularly polarized beam, passing only the orthogonally polarized diffracted beam. This allows readout of the grating as it is being recorded. The diffracted light is imaged onto the detector D. This system classifies input patterns presented to it into two classes according to whether the output of the detector when the pattern is presented is high or low. If during training a pattern we would like to classify as high yields a low response, the hologram is reinforced by exposing the crystal to the interference of the two beams, each carrying the image of that pattern. This exposure continues until the diffracted output increases by a fixed amount. If a pattern which should be classified as low is found during training to yield a diffracted output that is too high, the hologram diffracting that pattern is erased by a fixed amount by exposing the crystal with only one of the imaging beams. (One beam is blocked by the shutter SH). An experimental learning curve showing the diffracted intensities for each learning cycle for four training patterns in a system implemented using an Fe-doped $LiNbO_3$ crystal is shown in Fig. 6. The system classifies the patterns 0 and 2 as high and 1 and 3 as low. At first all patterns are low. The first two learning cycles are intended to drive the outputs of 0 and 2 above

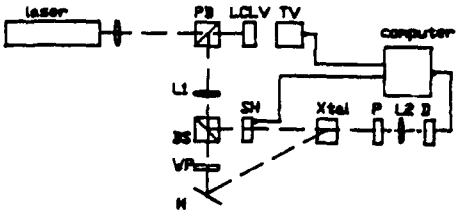


Fig. 5. Simple photorefractive learning system: *PB* is a polarizing beam splitter; *L1* and *L2* are imaging lenses; *WP* is a quarterwave plate; *SH* is a shutter; *P* is a polarizer; *D* is a detector; *M* is a mirror.

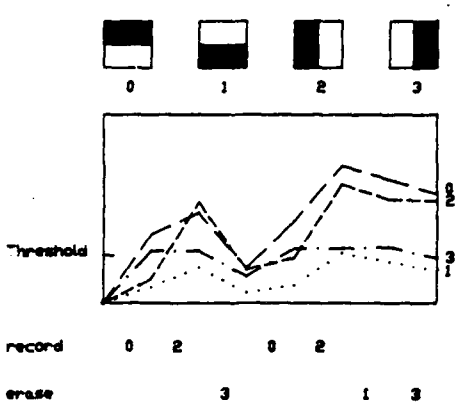


Fig. 6. Experimental learning curves.

threshold. However, they have the undesired effect of also driving pattern 3 above threshold. Thus in the third learning cycle 3 is erased. In this particular erase cycle the erasure was too severe. Note that pattern 2 is erased in this cycle, even though there is no overlap between this pattern and pattern 3. The reason for this is that the two images of pattern 3 are in focus only over a limited region of the crystal volume. Outside of this region the unfocused image may erase the hologram formed by pattern 2. In the subsequent two cycles patterns 0 and 2 are again reinforced. This has the unwanted effect of driving both patterns 1 and 3 just above threshold. In the final two cycles patterns 1 and 3 are erased until both are below threshold. At this point all patterns are correctly classified and learning stops.

In this experiment the photorefractive crystal acts as a 2-D modulator. The diffraction efficiency between the two imaging paths is high where the patterns 0 and 2 overlap and low where patterns 3 and 1 overlap. As mentioned above, a problem arises in the fact that the overlap is well defined only in the image plane, meaning the crystal must be thinner than the depth of focus of the images. To utilize the full capacity of photorefractive volume holograms it will be necessary to move beyond this implementation to architectures utilizing the full 3-D capacity of the crystal as discussed above. Nevertheless, this experiment demonstrates in a rudimentary way how learning in photorefractive crystals may proceed.

V. Conclusion

Photorefractive crystals represent a promising interconnection technology for optical neural computers. The ease of dynamic holographic modification of interconnections in these crystals allows the implementation of a large class of outer product learning networks. The density of interconnections which may be implemented in these crystals is limited by physical and geometrical constraints to the range of from 10^8 to 10^{10} per cm^3. To achieve these limits consideration must be given to the exposure schedule of the crystal.

The authors thank Xiang Guang Gu, Jeff Yu, and Hyuk Lee for many useful discussions relevant to the topics covered in this paper.

This research is supported by the Defense Advanced Research Projects Agency, the Air Force Office of Scientific Research, and the Army Research Office. David Brady acknowledges the support of the Office of Naval Research through the ONR/ASEE fellowship program.

Portions of this paper were presented at the IEEE International Conference on Neural Networks in San Diego, 21–24 June 1987.

References

1. Y. S. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," Sci. Am. **256**, 88 (1987).
2. D. Psaltis, J. Yu, X. G. Gu, and H. Lee, "Optical Neural Nets Implemented with Volume Holograms," in *Technical Digest of Topical Meeting on Optical Computing* (Optical Society of America, Washington, DC, 1987).
3. D. Psaltis, X. G. Gu, H. Lee, and J. Yu, "Optical Interconnections Implemented with Volume Holograms," to be published.
4. P. J. van Heerden, "Theory of Optical Information Storage in Solids," Appl. Opt. **2**, 393 (1963).
5. M. Cohen, "Design of a New Medium for Volume Holographic Information Processing," Appl. Opt. **25**, 2288 (1986).
6. K. Wagner and D. Psaltis, "Multilayer Optical Learning Networks," Proc. Soc. Photo-Opt. Instrum. Eng. **752**, 16 (1987).
7. K. Wagner and D. Psaltis, "Nonlinear Etalons in Adaptive Optical Neural Computers," presented at IEEE First Annual International Conference on Neural Networks, San Diego, 21–24 June 1987.
8. K. Wagner and D. Psaltis, "Multilayer Optical Learning Networks," Appl. Opt. **26**, 5061 (1987).
9. D. Z. Anderson, "Adaptable Interconnects for Optical Neuromorphs: Demonstration of a Photorefractive Projection Operator," in *Proceedings, International Conference on Neural Networks, San Diego* (June 1987).
10. T. Kohonen, *Self-Organization and Associative Memory* (Springer-Verlag, Berlin, 1984).
11. J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
12. S. S. Venkatesh and D. Psaltis, "Information Storage and Retrieval in Two Associative Nets," presented at Conference on Neural Network Models for Computing, Santa Barbara, CA (April 1985).
13. L. Personnaz, I. Guyon, and G. Dreyfus, "Information Storage and Retrieval in Spin-Glass Like Neural Networks," J. Phys. Lett. **46**, L359 (1985).
14. D. Psaltis and C. Park, "Nonlinear Discriminant Functions and Associative Memories," APS Conf. Proc. **151**, 370 (1986).
15. T. Maxwell, C. L. Giles, Y. C. Lee, and H. H. Chen, "Nonlinear Dynamics of Artificial Neural Systems," APS Conf. Proc. **151**, 299 (1986).
16. E. B. Baum, "On the Capabilities of Multilayer Perceptrons," to be published.
17. D. Psaltis and N. H. Farhat, "Optical Information Processing Based on an Associative Memory Model of Neural Nets with Thresholding and Feedback," Opt. Lett. **10**, 98 (1985).
18. Y. Owechko, G. J. Dunning, E. Marom, and B. H. Soffer, "Holographic Associative Memory with Nonlinearities in the Correlation Domain," Appl. Opt. **26**, 1900 (1987).
19. B. Kosko and C. Guest, "Optical Bidirectional Associative Memories," Proc. Soc. Photo-Opt. Instrum. Eng. **758**, (1987).
20. R. A. Athale, H. H. Szu, and C. B. Friedlander, "Optical Implementation of Associative Memory with Controlled Nonlinearity in the Correlation Domain," Opt. Lett. **11**, 482 (1986).
21. F. Rosenblatt, *Principles of Neurodynamics: Perceptron and the Theory of Brain Mechanisms* (Spartan Books, Washington, DC, 1961).
22. B. Widrow and M. E. Hoff, "Adaptive Switching Circuits," IRE WESCON Conv. Rec. **4**, 96 (1960).
23. D. E. Rumelhart and J. L. McClelland, Eds., *Parallel Distributed Processing, Vol. 1* (MIT Press, Cambridge, MA, 1986).
24. D. B. Parker, "Learning Logic," Invention Report S81-64, File 1, Office of Technology Licensing, Stanford U. (Oct. 1982).
25. J. D. Denker, Ed., "Neural Networks for Computing," APS Conf. Proc. 151 (1986).
26. A. D. Fisher, R. C. Fukuda, and J. N. Lee, "Implementations of Adaptive Associative Optical Computing Elements," Proc. Soc. Photo-Opt. Instrum. Eng. **625**, 196 (1986).
27. K. Fukushima, "A Hierarchical Neural Network Model for Associative Memory," Biol. Cybern. **50**, 105 (1984).
28. S. Grossberg, *Studies of Mind and Brain* (Reidel, Boston, 1982).
29. N. V. Kuktarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic Storage in Electrooptic Crystals. I: Steady State," Ferroelectrics **22**, 949 (1979).
30. J. W. Goodman, "Fan-In and Fan-Out with Optical Interconnections," Opt. Acta **32**, 1489 (1985).
31. D. L. Staebler and W. Phillips, "Fe-Doped $LiNbO_3$ for Read-Write Applications," Appl. Opt. **13**, 788 (1974).

BINARY FILTERS FOR PATTERN CLASSIFICATION

Santosh S. Venkatesh* and Demetri Psaltis
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

Revised 20 June 1988

Abstract

Generalised Fourier correlators imposing finite system space-bandwidth products are described and a class of binary filters is proposed. In pattern classification and signal registration applications it is shown that for a class of signals the binary filters yield the same asymptotic performance as the matched filter. It is hence adduced that a dynamic range of a single bit in the filter suffices for classification purposes. The effects of statistical side-lobe fluctuations and a finite system space-bandwidth product are included in the analysis. It is demonstrated that performance improves in a natural fashion with increase in the system space-bandwidth product for both the binary filter and the matched filter.

1 INTRODUCTION

Matched filters are commonly used in diverse applications in communication systems, signal processing, and pattern classification, where the task is typically the recognition of a particular signal or pattern immersed in noise. The principal theoretical argument supporting the use of Matched Filters is the classical result: Among the class of all linear filters, matched filters maximise a (suitably defined) signal- to-noise ratio [1]. Practical implementations of matched filters—and linear, shift-invariant systems, in general—are much facilitated by the fundamental Fourier convolution theorem wherein convolutions (or correlations) in one domain are transformed into products in the Fourier domain. As a consequence, relatively simple analog implementations such as optical Fourier-plane correlators [2], and digital implementations using algorithms such as the Fast Fourier Transform [3] abound.

The implementation of the system transfer function for the matched filter, however, requires a large dynamic range. A question of considerable theoretical and practical import is the determination of minimal complexity filters which have minimal dynamic range requirements, and for which good classification performance still attains (vis-à-vis the matched filter). The issue here is to determine the critical information needed for classification, and to discard redundant information. In this paper we propose a class of low complexity binary filters which are a step toward the resolution of this question. These filters encode information in the phase of the Fourier transform of the desired signal and require a dynamic range of just one bit.

*Currently with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104

Our principal theoretical result concerning the binary filters is the following: *For statistically uncorrelated pattern classes the binary filters provide the same asymptotic classification performance as the matched filter.* In fact, the binary filters provide classification performance comparable to (though bounded above by) the matched filter over all ranges.

These binary filters are of considerable practical importance. The requirement of a large dynamic range for the filter (corresponding to the many bits required to represent each sample point) is obviated, and just a single representation bit is utilised per sample point. The resultant decrease in required memory storage paves the way for low cost, low complexity systems—both digital and analog—which retain good classification performance. Of particular interest in optical filter implementations is the recent availability of a two-dimensional binary spatial light modulator—the magneto-optic device. We have demonstrated good classification in experimental optical correlators with our binary filters implemented using these devices [4].

In the next section we define a general family of bounded space-bandwidth product Fourier correlators, and formally prescribe the matched filter and the binary filter in this context. We also outline the signal statistics that we utilise, and set up a performance measure which incorporates information about both the correlation peak, and the side-lobe energy for all the pattern classes. In section 3 we analyse the performance of the matched filter and the binary filter in a two-class pattern recognition problem where the patterns belong to well-defined statistical classes, and are noise-free. We obtain analytical results for the performance measure as a function of the system space-bandwidth product in the two cases. In section 4 we investigate the attrition in classification performance in both systems when the input patterns are corrupted by additive noise. Sections 5 and 6 are devoted to numerical solutions and discussions of the comparative classification performance of the matched filter and the proposed binary filter: We demonstrate the monotonic improvement in performance in both systems as the system space-bandwidth product is increased, and show the asymptotic merging of the performance curves for the binary filter and the matched filter.

Notation: Let ω be some fixed (but arbitrary) positive quantity. To each real-valued function, f, of a real variable we associate its *finite-domain Fourier transform* F_ω formally defined by

$$F_\omega(u) = \int_{-\omega}^{\omega} f(x)\, e^{-i2\pi ux}\, dx\,. \tag{1}$$

We will use the terminology "space" for the variable x—the domain of the input signals—and "frequency" for the variable u—the domain of the associated Fourier transform.

2 FOURIER CORRELATORS

2.1 Bounded Space-Bandwidth Systems

The conventional Fourier correlator of equation correlator is shift invariant and admits signals of infinite space-bandwidth product (SBP) without loss of information. In this paper we will analyse the effect on classification performance of imposing a *finite system space-bandwidth product.* In particular, we consider shift variant Fourier correlators which process inputs through windows $(-\omega\,,\,\omega)$ in space, and $(-\nu\,,\,\nu)$ in frequency: For a given signal, $f(x)$, and reference, $h(x)$, the output, $g(x)$, of the bounded space-bandwidth correlator is given by

$$g(x) = \int_{-\nu}^{\nu} F_\omega(u)\, \bar{H}_\omega(u)\, e^{i2\pi ux}\, du\,.$$

We define the *system space-bandwidth product*, which we denote by p, to be the product of the width of the spatial and frequency windows: $p = 4\omega\nu$.

We consider two representative pattern classes, C_1 and C_2. The input signals are real valued functions, $f(x)$, which are sample realisations (drawn from some underlying probability distribution) of one of the two pattern classes . We will denote by $f_j(x)$ the input conditioned upon being drawn from pattern class C_j. For fixed system space-bandwidth product, p, we compare the following two classifiers for different choices of reference signal, $h(x)$.

Matched Filter: The reference signal, $h(x)$, is chosen matched to the sample realisation, $f_1(x)$, of class C_1. The correlation output, $g_j^M(x)$, for the matched filter conditioned upon class C_j at the input is given by

$$g_j^M(x) = \int_{-\nu}^{\nu} F_{\omega,j}(u)\, \bar{F}_{\omega,1}(u)\, e^{i2\pi ux}\, du\,. \tag{2}$$

If $\omega = \nu = \infty$, we have the classical matched filter. For finite p a correlation peak is still produced for class C_1. (Classification performance, however, deteriorates as p decreases.) Note that the matched filter above, in general, requires exponential dynamic range.

Binary Filter: The reference signal, $h(x)$, is chosen such that

$$H_\omega(u) = \mathrm{sgn}\{\Re\{F_{\omega,1}(u)\}\} = \begin{cases} 1 & \text{if } \Re\{F_{\omega,1}(u)\} \geq 0 \\ -1 & \text{if } \Re\{F_{\omega,1}(u)\} < 0 \end{cases}.$$

The filter hence takes on values -1 and +1 only at each frequency, so that we have a dynamic range of one bit. The correlation output, $g_j^B(x)$, of the binary filter conditioned upon class C_j at the input is given by

$$g_j^B(x) = \int_{-\nu}^{\nu} F_{\omega,j}(u)\, \mathrm{sgn}\{\Re\{F_{\omega,1}(u)\}\}\, e^{i2\pi ux}\, du\,. \tag{3}$$

Note that the binary filter tracks the phase of $F_{\omega,1}(u)$, so that we can expect a correlation peak for class C_1, but not for class C_2.

In figure 1 we demonstrate two correlations of a random one-dimensional input sequence; in figure 1(a) the correlation was accomplished using a matched filter, while in figure 1(b) the correlation was performed using a binary filter. As seen, the correlation peaks and side-lobe fluctuation levels are essentially indistinguishable in the two cases.

2.2 Performance Measure

In characterising the classification performance of the two filters, we concentrate on two key measures: The strength of the correlation peak, and the side-lobe structure. For specific sample realisations not much can be said about the size of the side-lobes; however, if signal statistics are known we can extract peak and side-lobe information from a consideration of the ensemble. In the next section we describe a specific statistical structure for the two signal classes from which we can obtain quantitative estimates of filter performance.

For $j = 1, 2$ let $g_j(x)$ denote filter output conditioned upon class C_j being present at the input. Define

$$\mu_j = \sup_x\{|\mathbf{E}\{g_j(x)\}|\}\,,$$
$$\eta_j = \sup_x\{\mathrm{Var}\{g_j(x)\}\}\,.$$

We define the *performance coefficient* ρ by

$$\rho = \frac{(\mu_1 - \mu_2)^2}{\eta_1 + \eta_2} . \tag{4}$$

The term in the numerator measures the relative size of correlation peaks for the two classes, while the term in the denominator factors in the average energy in the side-lobes. The coefficient, ρ, hence is an indicator of how well the filter discriminates class C_1 from class C_2.

We denote by ρ^M and ρ^B, respectively, the performance coefficient for the matched filter and the binary filter. We shall take system performance to be a monotonically increasing function of the coefficient ρ, with the system with the largest ρ realising the best performance.

Note that the form of the coefficient ρ is similar to a *signal- to-noise ratio*, the "signal" corresponding to class C_1 and the "noise" to class C_2. (In fact, when the output variable $g(x)$ is Gaussian, and the *a priori* probabilities of the two classes are the same, it turns out that the form of the Bhattacharyya coefficient [5] is identical to equation 4 for ρ). From classical communication theory we have that for correlational-systems which are linear functionals of the input signal, the peak signal-to-noise ratio for a signal immersed in white noise is obtained for the matched filter. Hence we expect the classification performance of the binary filter to be bounded by that of the matched filter.

2.3 Signal Statistics

In order to facilitate analysis we assume a specific statistical structure for the ensemble os signals in the two classes. We assume that the signals $f_1(x)$ and $f_2(x)$ corresponding to the two classes C_1 and C_2 are sample realisations of mutually independent, white random processes with

$$\begin{aligned} \mathbf{E}\{f_j(x)\} &= 0 , \\ \mathbf{E}\{f_j(x) f_j(y)\} &= \sigma_j^2 \delta(x-y) . \end{aligned} \tag{5}$$

The signal classes have been restricted to be stationary and white in order to effect some simplicity in the ensuing analysis. The stationarity constraint can be relaxed to allow of correlation functions of the form $r_j(x)\delta(x-y)$; the analysis for this case is essentially the same as for the case we consider. With the added constraint that the process be Gaussian, one or both constraints can be relaxed to encompass general correlation functions of the form $r_j(x,y)$.

From equation 1 the real and imaginary parts of $F_{\omega,j}(u)$ are given by

$$\begin{aligned} \Re\{F_{\omega,j}(u)\} &= \int_{-\omega}^{\omega} f_j(x) \cos 2\pi u x \, dx , \\ \Im\{F_{\omega,j}(u)\} &= \int_{-\omega}^{\omega} f_j(x) \sin 2\pi u x \, dx . \end{aligned} \tag{6}$$

The random processes $f_j(x)$ are independent and zero mean. By virtue of the *Central Limit Theorem* then, it can be readily seen to follow that $\Re\{F_{\omega,1}(u)\}$, $\Im\{F_{\omega,1}(u)\}$, $\Re\{F_{\omega,2}(u)\}$, and $\Im\{F_{\omega,2}(u)\}$ are mutually independent Gaussian random processes with zero mean. Some algebraic manipulation readily yields the following:

$$\mathbf{E}\{\Re\{F_{\omega,j}(u)\}\Im\{F_{\omega,j}(t)\}\} = 0 , \tag{7}$$

$$\mathbf{E}\{\Re\{F_{\omega,j}(u)\}\Re\{F_{\omega,j}(t)\}\} = \sigma_j^2 \omega [\operatorname{sinc} 2\omega(u-t) + \operatorname{sinc} 2\omega(u+t)] , \tag{8}$$

$$\mathbf{E}\{\Im\{F_{\omega,j}(u)\}\Im\{F_{\omega,j}(t)\}\} = \sigma_j^2 \omega [\operatorname{sinc} 2\omega(u-t) - \operatorname{sinc} 2\omega(u+t)] . \tag{9}$$

We also require the first and second moments of the random processes $\mathrm{sgn}\{\Re\{F_{\omega,j}(u)\}\}$ and $|\Re\{F_{\omega,j}(u)\}|$. Define $r_\omega : \mathbb{R}^2 \to [-1,1]$ by

$$r_\omega(u,t) = \frac{\mathrm{sinc}\,2\omega(u-t) + \mathrm{sinc}\,2\omega(u+t)}{(1+\mathrm{sinc}\,4\omega u)^{1/2}(1+\mathrm{sinc}\,4\omega t)^{1/2}}\,. \tag{10}$$

Note that from equation 8 it follows that for each u and t, $r_\omega(u,t)$ is just the correlation coefficient of the random variables $\Re\{F_{\omega,j}(u)\}$ and $\Re\{F_{\omega,j}(t)\}$. The following results can be readily shown (cf. [6], for instance).

$$\mathrm{E}[\mathrm{sgn}\{\Re\{F_{\omega,j}(u)\}\}] = 0\,, \tag{11}$$

$$\mathrm{E}\{|\Re\{F_{\omega,j}(u)\}|\} = \frac{\sqrt{2\omega\sigma_j^2}}{\sqrt{\pi}}\sqrt{1+\mathrm{sinc}\,4\omega u}\,, \tag{12}$$

$$\mathrm{E}[\mathrm{sgn}\{\Re\{F_{\omega,j}(u)\}\}\,\mathrm{sgn}\{\Re\{F_{\omega,j}(t)\}\}] = \frac{2}{\pi}\sin^{-1} r_\omega(u,t)\,, \tag{13}$$

$$\begin{aligned}\mathrm{E}\{|\Re\{F_{\omega,j}(u)\}|\,|\Re\{F_{\omega,j}(t)\}|\} &= \frac{2\omega\sigma_j^2}{\pi}\Big[(1+\mathrm{sinc}\,4\omega u)^{1/2}(1+\mathrm{sinc}\,4\omega t)^{1/2}(1-r_\omega(u,t)^2)^{1/2} \\ &+ (\mathrm{sinc}\,2\omega(u-t)+\mathrm{sinc}\,2\omega(u+t))\,\sin^{-1} r_\omega(u,t)\Big]\,.\end{aligned} \tag{14}$$

3 TWO-CLASS DISCRIMINATION

3.1 The Matched Filter

Our consideration of the matched filter as a correlational system described by equation 2 differs somewhat from the classical deterministic matched filter [1] in the inclusion of a finite system space-bandwidth and the representation of both input and reference signals as members of a statistical class. The performance coefficient that we derive hence reflects the relative correlation peaks, and the "noisy" side-lobe fluctuations averaged over the ensemble as a function of p (the system SBP).

We estimate the parameters, μ_j and η_j, in equation 4 in turn for the two classes using the results tabulated in section 2.3.

Class C_1: The system output is given by

$$g_1^M(x) = \int_{-\nu}^{\nu}\left([\Re\{F_{\omega,1}(u)\}]^2 + [\Im\{F_{\omega,1}(u)\}]^2\right)e^{i2\pi ux}\,du\,.$$

A simple computation yields:

$$\mu_1 = 4\omega\nu\sigma_1^2\,, \tag{15}$$

$$\eta_1 = 64\omega^2\nu^2\sigma_1^4\int_0^1(1-t)(\mathrm{sinc}\,4\omega\nu t)^2\,dt\,. \tag{16}$$

Class C_2: The system output is given by

$$g_2^M(x) = \int_{-\nu}^{\nu} F_{\omega,2}(u)\, \bar{F}_{\omega,1}(u)\, e^{i2\pi ux}\, du\,. \tag{17}$$

The correlation peak and average side-lobe energy can again be simply estimated:

$$\mu_2 = 0\,. \tag{18}$$

$$\eta_2 = 32\omega^2\nu^2\sigma_1^4 \int_0^1 (1-t)(\mathrm{sinc}\, 4\omega\nu t)^2\, dt\,. \tag{19}$$

Defining α as a function of the space-bandwidth product p by

$$\alpha(p) = \int_0^1 (1-t)(\mathrm{sinc}\, pt)^2\, dt\,, \tag{20}$$

the performance coefficient of equation 4 is hence given by

$$\rho^M = \frac{\sigma_1^2/\sigma_2^2}{2\alpha(p)\,[1 + 2\sigma_1^2/\sigma_2^2]}\,. \tag{21}$$

Asymptotic results: The above expression can be readily evaluated for extreme values of the system space-bandwidth product. For very low space-bandwidth products $\alpha(p)$ approaches $1/2$, so that

$$\rho^M \longrightarrow \frac{\sigma_1^2/\sigma_2^2}{1 + 2\sigma_1^2/\sigma_2^2} \qquad \text{as } p \to 0\,.$$

For very high space-bandwidth products, on the other hand, $\alpha(p)$ asymptotically approaches the value $1/2p$, so that

$$\rho^M \longrightarrow \frac{p\sigma_1^2/\sigma_2^2}{1 + 2\sigma_1^2/\sigma_2^2} \qquad \text{as } p \to \infty\,.$$

The asymptotic results correspond well with intuition. For very low space- bandwidth products we expect a low processing gain for the system as not much correlation matching can be obtained. For high space-bandwidth products on the other hand, the use of uncorrelated signals at the input yields large processing gains increasing linearly with the space-bandwidth product.

It is instructive to compare the performance measure given by equation 21 with the classical matched filter result for the signal-to-noise ratio (SNR) of a deterministic signal immersed in white noise. The *processing gain* of a classical system (defined to be the ratio of the output SNR to the input SNR) is given essentially by the signal space-bandwidth product [1]. If we define σ_1^2/σ_2^2 to be a measure of the input SNR for the statistical case under consideration, then the processing gain of our system, in the limit of large p and small input SNR, is given by $1/2\alpha(p) \approx p$, which is precisely the classical result. (The additional input SNR dependent term present in the denominator of equation 21 arises because the statistical side-lobe fluctuations are also taken into account in our performance measure; this term will not be significant for low input SNR scenarios.) In fine, the presence of a finite system space-bandwidth product manifests itself in a loss of processing gain; the larger the space-bandwidth product, the more the processing gain realised by the system.

3.2 The Binary Filter

The system output conditioned upon class C_j being present at the input is given by equation 3. We again estimate the parameters, μ_j and σ_j, for the two classes in turn.

Class C_1: The output of the system with $f_1(x)$ at the input is given by substitution in equation 3:

$$g_1^B(x) = \int_{-\nu}^{\nu} |\Re\{F_{\omega,1}(u)\}|\, e^{i2\pi ux}\, du + i \int_{-\nu}^{\nu} \Im\{F_{\omega,1}(u)\} \operatorname{sgn}\{\Re\{F_{\omega,1}(u)\}\}\, e^{i2\pi ux}\, du\,. \quad (22)$$

From equations 7 and 12 it then follows that

$$\mu_1 = \sup_x \{|\mathrm{E}\{g_1^B(x)\}|\} = \frac{\sqrt{8\omega\nu^2\sigma_1^2}}{\sqrt{\pi}} \int_0^1 (1 + \operatorname{sinc} pt)^{1/2}\, dt\,, \quad (23)$$

where, as before, p is the space-bandwidth product $4\omega\nu$.

Now, in equation 22 set

$$\begin{aligned} k_1(x) &= \int_{-\nu}^{\nu} |\Re\{F_{\omega,1}(u)\}|\, e^{i2\pi ux}\, du\,, \\ k_2(x) &= \int_{-\nu}^{\nu} \Im\{F_{\omega,1}(u)\} \operatorname{sgn}\{\Re\{F_{\omega,1}(u)\}\}\, e^{i2\pi ux}\, du\,. \end{aligned}$$

Then

$$g_1^B(x) = k_1(x) + ik_2(x)\,.$$

Clearly, $k_1(x)$ and $k_2(x)$ are uncorrelated complex random processes with $k_2(x)$ being zero mean. Hence

$$\operatorname{Var}\{g_1^B(x)\} = \mathrm{E}\{|k_1(x)|^2\} + \mathrm{E}\{|k_2(x)|^2\} - |\mathrm{E}\{k_1(x)\}|^2\,.$$

Using equation 9 and equations 12–14 we obtain after some algebraic manipulation that

$$\begin{aligned} \operatorname{Var}\{g_1^B(x)\} &= \frac{4\omega\nu^2\sigma_1^2}{\pi} \int_{-1}^{1}\int_{-1}^{1} \Big[\operatorname{sinc}\{\tfrac{p}{2}(u-t)\}\, \sin^{-1} r_{p/4}(u,t) - \frac{1}{2}(1+\operatorname{sinc} pu)^{1/2} \\ &\times\ (1+\operatorname{sinc} pt)^{1/2}(1-\sqrt{1-r_{p/4}(u,t)^2})\Big] \cos 2\pi(u-t)\nu x\, du\, dt\,. \end{aligned} \quad (24)$$

Note that $r_\omega(\nu u, \nu t) = r_{p/4}(u,t)$, which can be verified by direct substitution in the defining equation 10 with $p = 4\omega\nu$.

No analytic expression is available in general for $\eta_1 = \sup_x \{\operatorname{Var} g_1^B(x)\}$, and we have to resort to numerical evaluation for specified parameters p, σ_1^2, and σ_2^2. (Note that in general, the supremum does not occur at $x = 0$.)

Class C_2: From equation 3, the output for class C_2 is given by

$$g_2^B(x) = \int_{-\nu}^{\nu} F_{\omega,2}(u) \operatorname{sgn}\{\Re\{F_{\omega,1}(u)\}\}\, e^{i2\pi ux}\, du\,.$$

Again having recourse to section 2.3, we can show that

$$\mu_2 = \sup_x \{|\mathrm{E}\{g_2^B(x)\}|\} = 0\,, \quad (25)$$

$$\operatorname{Var}\{g_2^B(x)\} = \frac{4\omega\sigma_2^2}{\pi} \int_{-1}^{1}\int_{-1}^{1} \operatorname{sinc}\{\tfrac{p}{2}(u-t)\}\, \sin^{-1} r_{p/4}(u,t) \cos 2\pi(u-t)\nu x\, du\, dt\,. \quad (26)$$

Again, no analytic expression can be found for $\eta_2 = \sup_x \{\operatorname{Var} g_2^B(x)\}$, in general, and we must resort to numerical evaluation.

Define β_0, β_1, and β_2 as functions of the space- bandwidth product p by

$$\beta_0(p) = \left[\int_0^1 (1+\operatorname{sinc} pt)^{1/2}\,dt\right]^2 , \tag{27}$$

$$\beta_1(p) = \sup_x \int_{-1}^{1}\int_{-1}^{1} \operatorname{sinc}\{\tfrac{p}{2}(u-t)\}\,\sin^{-1} r_{p/4}(u,t)\,\cos 2\pi(u-t)\nu x\,du\,dt , \tag{28}$$

$$\beta_2(p) = \sup_x \int_{-1}^{1}\int_{-1}^{1} \Big\{\operatorname{sinc}\{\tfrac{p}{2}(u-t)\}\,\sin^{-1} r_{p/4}(u,t) - \tfrac{1}{2}(1+\operatorname{sinc} pu)^{1/2} \times (1+\operatorname{sinc} pt)^{1/2}\{1-(1-r_{p/4}(u,t)^2)^{1/2}\}\Big\}\,\cos 2\pi(u-t)\nu x\,du\,dt . \tag{29}$$

Combining the results of equations 23–26, and using the defining equations 27–29 we obtain the performance coefficient, ρ^B, of the binary filter to be

$$\rho^B = \frac{2\beta_0(p)\,\sigma_1^2/\sigma_2^2}{\beta_1(p)+\beta_2(p)\,\sigma_1^2/\sigma_2^2} . \tag{30}$$

We will return to a comparative analysis of the expressions 21 and 30 in section 5.

4 CLASSIFICATION IN ADDITIVE NOISE

In practice, the issue of *system robustness* in the face of signal degradations, and noise becomes important. We illustrate how noisy signals result in performance attrition in the two correlator systems.

We consider the case where the input signal $f(x)$ is contaminated by an additive noise term $n(x)$. (We assume that the reference signal, $h(x)$, being known *a priori* can hence be represented in a reasonably accurate and noise-free manner). We take $n(x)$ to be an independent noise process which is additive and white with

$$\begin{aligned} \mathbf{E}\{n(x)\} &= 0 , \\ \mathbf{E}\{n(x)\,n(y)\} &= \sigma_n^2\delta(x-y) . \end{aligned}$$

The input signal term is then $f_j(x)+n(x)$, and the reference signal term (matched to class C_1) is $f_1(x)$.

4.1 The Matched Filter

Let $g^{\mathrm{M}}_{j,n}(x)$ denote the (noisy) correlation output of the system when the input signal is a noisy realisation of class C_j, viz., $f_j(x)+n(x)$. Then

$$\begin{aligned} g^{\mathrm{M}}_{j,n}(x) &= \int_{-\nu}^{\nu} F_{\omega,j}(u)\,\bar{F}_{\omega,1}(u)\,e^{i2\pi ux}\,du + \int_{-\nu}^{\nu} N_\omega(u)\,\bar{F}_{\omega,1}(u)\,e^{i2\pi ux}\,du \\ &= g^{\mathrm{M}}_j(x) + g^{\mathrm{M}}_n , \end{aligned}$$

where the first term, $g^{\mathrm{M}}_j(x)$, is the noise-free system response of equation 2 and the second term, g^{M}_n, is the additive noise term in the output correlation. The noise term independent of the signal term, and is zero mean with peak variance at the origin

$$\mathrm{Var}\{g^{\mathrm{M}}_n(0)\} = 2p^2\sigma_1^2\sigma_n^2\alpha(p) ,$$

identical in form to equation 19. Hence, using equations 15–20 we have

$$\begin{aligned}
\mu_{j,n} &= \sup_x \{|\mathbf{E}\{g^M_{j,n}(x)\}|\} = \mu_j \,, \\
\eta_{1,n} &= \sup_x \{\mathrm{Var}\{g^M_{1,n}(x)\}\} = 2p^2\sigma_1^2\,\alpha(p)(2\sigma_1^2+\sigma_n^2) \,, \\
\eta_{2,n} &= \sup_x \{\mathrm{Var}\{g^M_{2,n}(x)\}\} = 2p^2\sigma_1^2\,\alpha(p)(\sigma_2^2+\sigma_n^2) \,.
\end{aligned}$$

The performance coefficient ρ^M_n for the matched filter when input noise is present is hence given by

$$\rho^M_n = \frac{(\mu_{1,n}-\mu_{2,n})^2}{\eta_{1,n}+\eta_{2,n}} = \frac{\left(\frac{\sigma_1^2}{\sigma_2^2+2\sigma_n^2}\right)}{2\alpha(p)\left(1+\frac{2\sigma_1^2}{\sigma_2^2+2\sigma_n^2}\right)} \tag{31}$$

where $\alpha(p)$ is as defined in equation 20.

A comparison of equations 21 and 31 shows that the presence of additive input noise is equivalent to an additive increase in the variance (or spread) of class C_2 by exactly twice the spread of the noise.

4.2 The Binary Filter

Tracing through an analogous analysis yields the performance coefficient ρ^B_n for the binary filter when the input is degraded by additive noise. In general, however, it turns out that the form of ρ^B_n is not conducive to a convenient representation as in equation 30 for the noise-free case; specifically, in equation 29, the functional $\beta_2(p)$ has to be replaced by a more complicated supremum taken over the sum of two integrals, the coefficient of one being σ_1^2, and of the other being σ_n^2. (The supremum is now a function of not only the space-bandwidth product p, but also of the signal and noise variances.) Using $\sup\{A+B\} \le \sup\{A\}+\sup\{B\}$, we can arrive at the following convenient lower bound estimate for ρ^B_n for the sake of comparison:

$$\rho^B_n \ge \frac{2\beta_0(p)\frac{\sigma_1^2}{\sigma_2^2+2\sigma_n^2}}{\beta_1(p)+\beta_2(p)\frac{\sigma_1^2}{\sigma_2^2+2\sigma_n^2}} \tag{32}$$

with the functionals $\beta_0(p)$, $\beta_1(p)$, and $\beta_2(p)$ given by equations 27–29.

On comparing equations 30 and 32 we see that the effect of additive noise is to create a larger effective spread for class C_2 just as in the case of the matched filter. In both cases, the noise effectively reduces the ability of the system to pick out class C_1 by increasing side-lobe energy, and at the same time increasing the correlation spread of class C_2.

5 NUMERICAL SOLUTIONS AND DISCUSSION

Let σ^2 denote the ratio $\sigma_1^2/\sigma_2^2+2\sigma_n^2$. We will refer to σ^2 as the *class spread ratio*; in essence σ^2 is astatistical measure of the relative strengths of "signal" (class C_1) and "noise"(class C_2, and additive noise) at the input of the correlational system. Recapitulating the expressions for the performance coefficients for easy reference, we have

$$\rho^M = \frac{\sigma^2}{2\,\alpha(p)+4\,\alpha(p)\,\sigma^2} \,,$$

$$\rho^B = \frac{2\beta_0(p)\sigma^2}{\beta_1(p)+\beta_2(p)\sigma^2},$$

where the functionals $\alpha(p)$, $\beta_0(p)$, $\beta_1(p)$, and $\beta_2(p)$ are defined, respectively, in equation 20 and equations 27–29.

A numerically generated family of performance curves for the two systems is depicted in figures 2 and 3. In each figure the performance coefficient, ρ, is plotted as a function of the class spread ratio, σ^2, and the family of curves is generated by varying the space-bandwidth parameter p between 8 and 256. In order to facilitate comparison between the matched filter and the binary filter, for values of $p = 8$, and $p = 256$, the corresponding performance curves of the two systems are extracted from figures 2 and 3, and plotted on the same graph in figures 4 and 5.

It can be immediately seen from the figures that, all other things being held constant, *the performance coefficient ρ is a monotonically increasing function of the system space-bandwidth product* for both filtration systems. This is clearly in accordance with our expectations as increasing the system space-bandwidth product is equivalent to increasing the size of the windows in the space and frequency domains, so that a greater degree of correlation matching can be obtained.

Now, when the class spread ratio, σ^2, is large, we have a situation where the noise power, σ_n^2, and the class C_2 spread, σ_2^2, are both much smaller than the class C_1 spread, σ_1^2. This can be viewed as essentially saying that patterns of class C_1 can take on values from a much wider set than can patterns of class C_2 and the noise patterns. The probability of significant cross-correlation in any particular case is then quite small, so that we expect good classification performance for large values of σ. This intuitive expectation is echoed in figures 2–5, where we see that for the matched filter and the binary filter, *the performance coefficient ρ is a monotonically increasing function of the class spread ratio, σ^2*, for each performance curve (corresponding to fixed p).

For the matched filter, a close examination of the asymptotes and the slope near the origin of each performance curve reveals that *"large p" behaviour holds for relatively small values of the system space-bandwidth product* (as small as p=8). The asymptote of the performance curve for the matched filter is approximately $p/2$, and the graph near the origin is a straight line with positive slope p.

Though ρ^B is always bounded from above by ρ^M, *for large class spread ratios the performance curve of the binary filter approaches the same asymptote, $p/2$, as the matched filter*, so that their performance is virtually identical. An examination of their relative performance for each p in the range considered indicates that when the class spread ratio is unity (i.e., the two classes have the same variance), we have $\rho^B \approx 2\rho^M/3$.

6 CONCLUSION

These numerical simulations, coupled with the prior success of experimental systems utilising binary filters [4], tend to bolster the intuitive notion that the phase of the Fourier Transform contains most of the information content in the signal. The significance of the results lies in the demonstration that, for classification purposes, most of the information content in the signal can be extracted with filters of low complexity. Specifically, the binary filters of this paper require only a single bit dynamic range but provide classification performance comparable to the matched filter which is much more prodiguous in its dynamic range requirements. While the success of these schemes is very encouraging, some questions remain: We have demonstrated binary correlator structures based on heuristic algorithms; however, it is not immediately obvious whether we can specify *optimum* binary correlator structures for a given problem. As a specific instance, we can obtain

filters which maximally separate pattern classes in that the filter is orthogonal to all unwanted patterns, while yielding a significant correlation only if the desired pattern is present. It is not clear, however, whether an algorithm can be specified which yields the binary filter which is the best approximation to any such maximally separating filter.

Acknowledgement

This research was supported by the Army Research Office and General Dynamics.

References

[1] G. L. Turin, "An introduction to matched filters," *IRE Trans. Inform. Theory*, vol. IT–6, pp. 311–329, 1960.

[2] A. Vander Lugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inform. Theory*, vol. IT–10, pp. 139–145, 1964.

[3] A. V. Oppenheim and R. V. Schafer, *Digital Signal Processing.* Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

[4] D. Psaltis, E. G. Paek, and S. S. Venkatesh,"Optical image correlation using a binary spatial light modulator," *Opt. Eng.*, vol. 23, pp. 698–704, 1984.

[5] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Comm. Tech.*, vol. COM–15, pp. 52–60, 1967.

[6] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT–4, pp. 69–72, 1958.

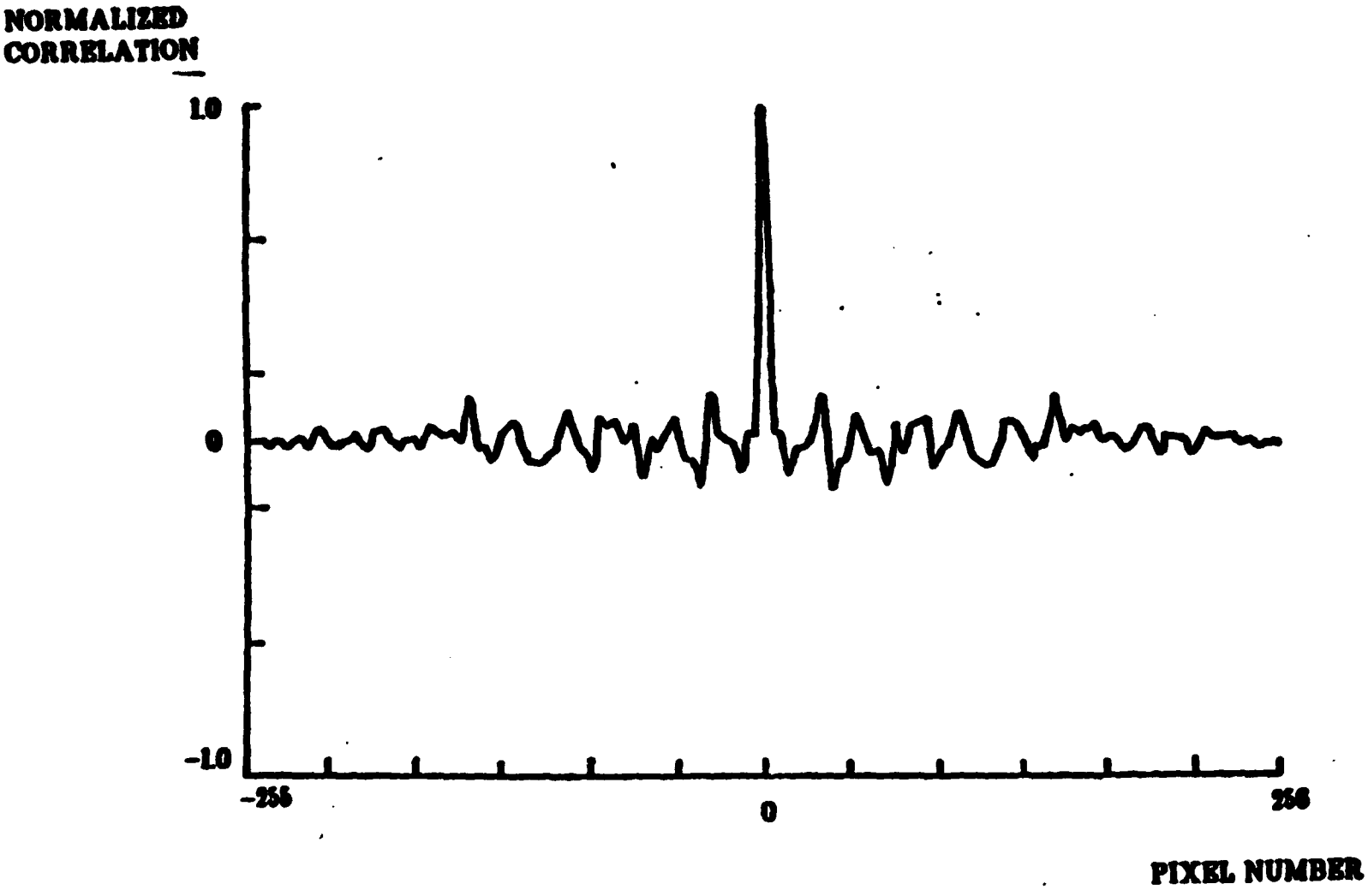


Fig. 1(a). Correlation of a random sequence using a Matched Filter [4].

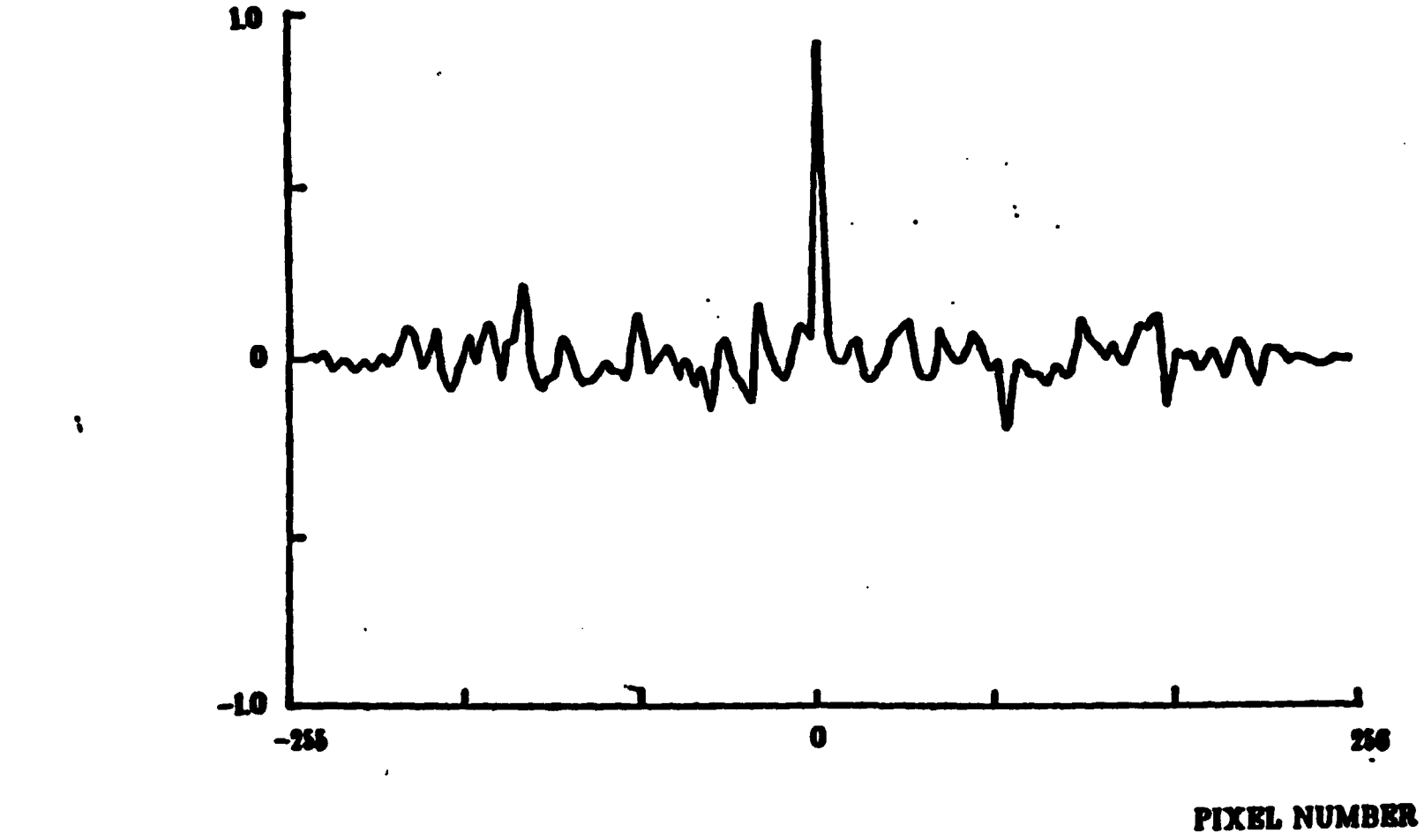


Fig. 1(b). Correlation of a random sequence using a Binary Filter [4].

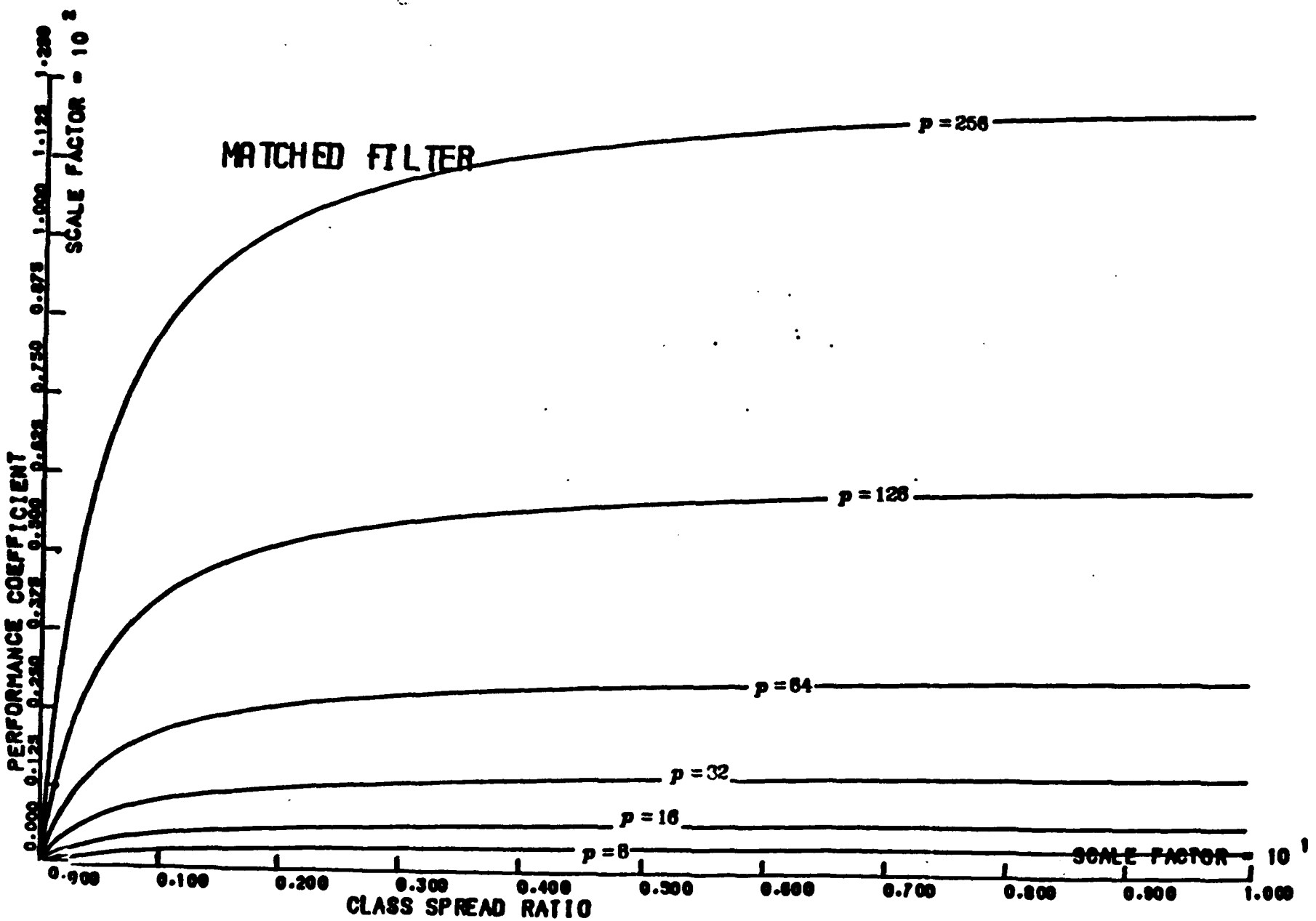


Fig. 2. Plot of the performance coefficient, ρ_m, of the Matched Filter *vs.* the class spread ratio, σ^2, with the system space-bandwidth product, p, as a parameter.

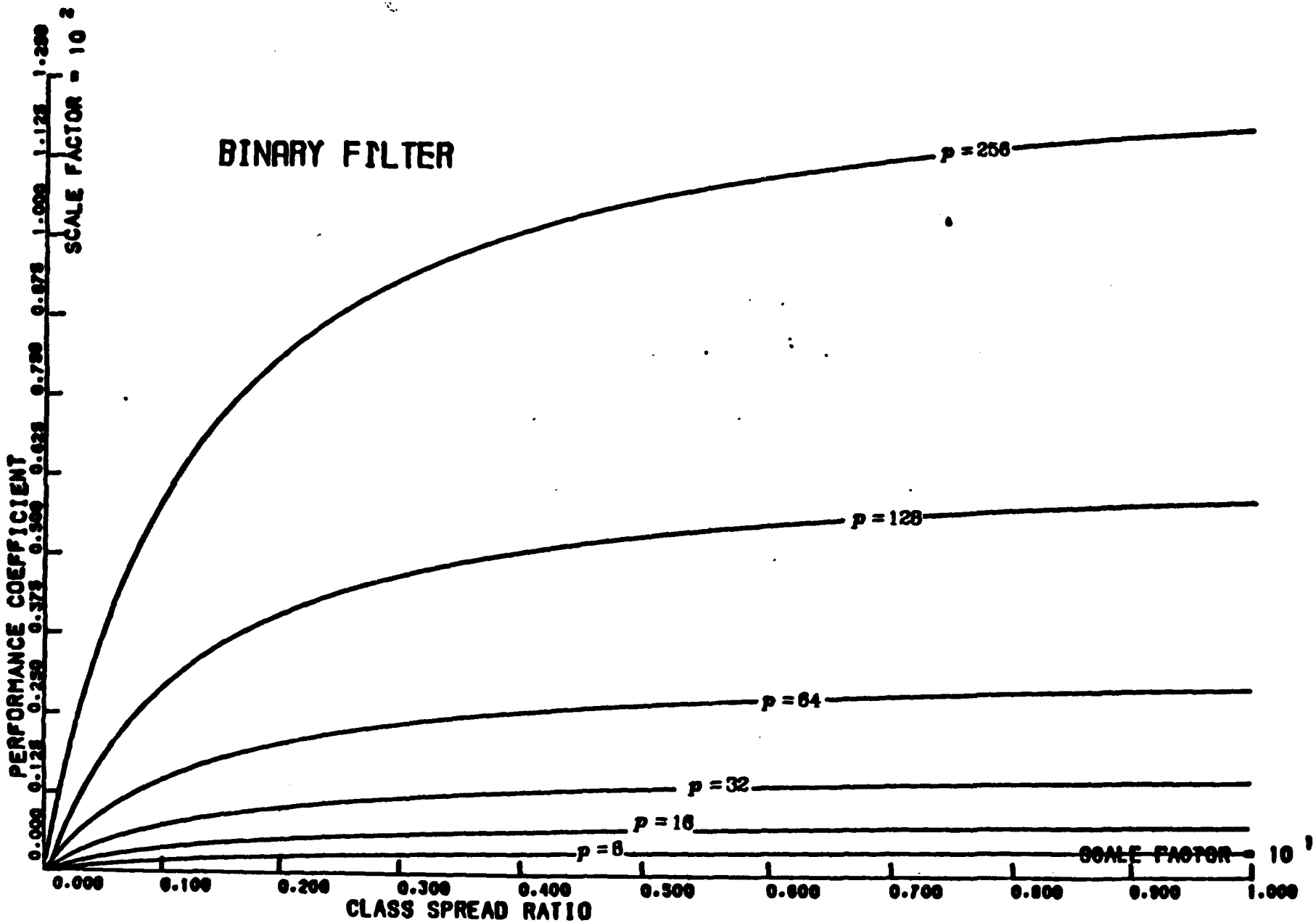


Fig. 3. Plot of the performance coefficient, ρ_b, of the Binary Filter *vs.* the class spread ratio, σ^2, with the system space-bandwidth product, p, as a parameter.

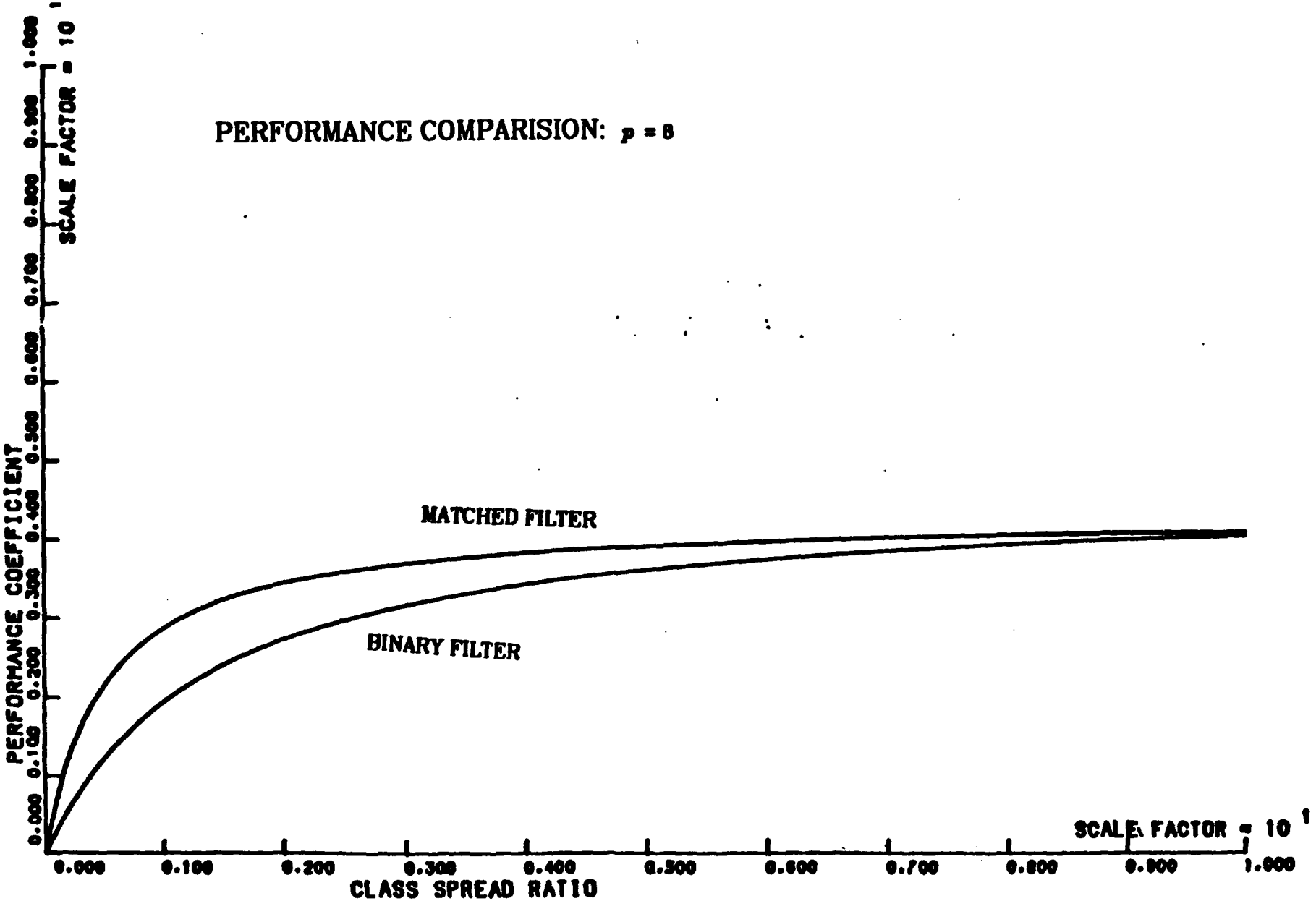


Fig. 4. Plots of the relative performance of the Matched Filter and the Binary Filter for a given system space-bandwidth product p = 8

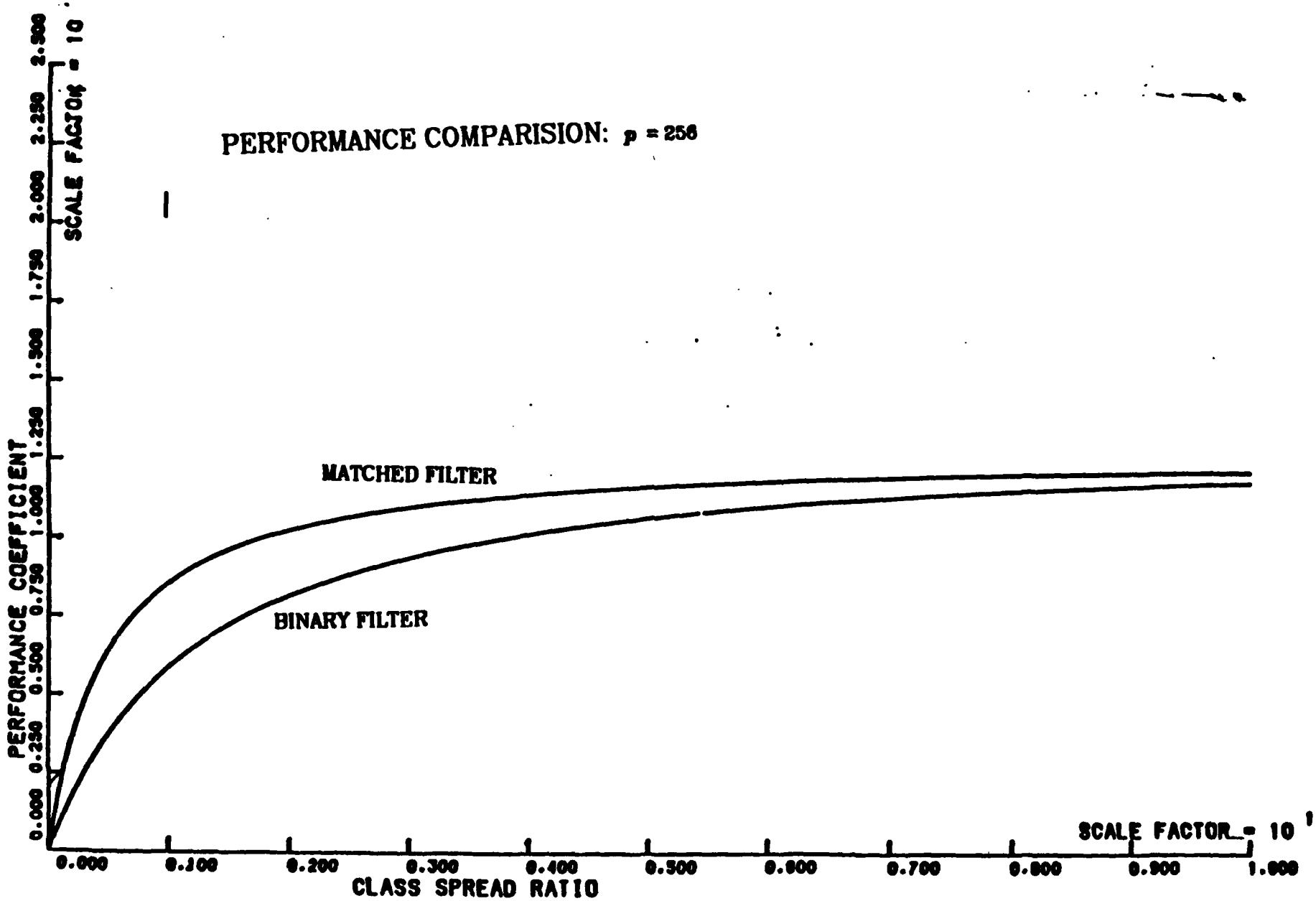


Fig. 5. Plots of the relative performance of the Matched Filter and the Binary Filter for a given system space-bandwidth product $p = 256$.

Neural Networks, Vol. 1, pp. 149–163, 1988
Printed in the USA.

0893-6080/88 $3.00 + .00

ORIGINAL CONTRIBUTION

Higher Order Associative Memories and Their Optical Implementations*

DEMETRI PSALTIS, CHEOL HOON PARK, AND JOHN HONG†

California Institute of Technology

(*Received and accepted* 22 *December* 1987)

Abstract—*The properties of higher order memories are described. The non-redundant, up to Nth order polynomial expansion of N-dimensional binary vectors is shown to yield orthogonal feature vectors. The properties of expansions that contain only a single order are investigated in detail and the use of the sum of outer product algorithm for training higher order memories is analyzed. Optical implementations of quadratic associative memories are described using volume holograms for the general case and planar holograms for shift invariant memories.*

1. INTRODUCTION

An associative memory can be thought of as a system that stores a prescribed set of vector pairs $(\mathbf{x}^m, \mathbf{y}^m)$ for $m = 1, \ldots, M$ and also produces $\mathbf{y}^m$ as its output when $\mathbf{x}^m$ becomes its input. We denote by N and N_0 the dimensionalities of the input and output vectors, respectively. When the output vectors are stored as binary N_0-tuples, the associative memory can be implemented as an array of discriminant functions, each dichotomizing the input vectors into two classes. This type of associative memory is shown schematically in Figure 1. In evaluating the effectiveness of a particular associative memory we are concerned with its ability to store a large number of associations (capacity), the ease with which the parameters of the memory can be set to realize the prescribed mappings (learning), and how it responds to inputs that are not members of its training set (generalization). In this paper we discuss a class of associative memories known as higher order memories that have been recently investigated by a number of separate research groups (Baldi & Venkatesh, 1987; Chen *et al.*, 1986; Giles & Maxwell, 1987; Maxwell, Giles, Lee, & Chen, 1986; Newman, 1987; Poggio, 1975; Psaltis & Park, 1986; Sejnowski, 1986). Our motivation for investigating these memories was the increase in storage capacity that results from the increase in the number of independent parameters or degrees of freedom that is needed to describe a higher order associative mapping. The relationship between the degrees of freedom of a memory and its ability to store associations (Abu-Mostafa & Psaltis, 1985) is fundamental to this work and we state it in the following subsection as a theorem.

1.1 Degrees of Freedom and Storage Capacity

Let D be the number of independent variables (degrees of freedom) we have under our control to specify input–output mappings and let each parameter have K separate levels or values that it can assume. We define the storage capacity C to be the maximum number of arbitrary associations that can be stored and recalled without error.

Theorem 1.

$$C \leq \frac{D \log_2 K}{N_0}. \tag{1}$$

Proof: The number of different states of memory is given by K^D and the total number of outputs that a given set of M input patterns can be mapped to is $2^{N_0 M}$. If the number of mappings were larger than the number of distinct states of the memory, then mappings would exist that are not implementable. Requiring that all mappings can be done leads to the relationship of the theorem.

The equality in (1) is achieved by Boolean circuits such as programmable logic arrays and an extreme case of a higher order memory we will discuss later. When the equality holds, resetting any one bit in any one of

* Funded by the Air Force Office of Scientific Research, the Army Research Office and the Defense Advanced Research Projects Agency.

† Dr. Hong is now with the Rockwell Science Center, Thousand Oaks, CA 91360.

Requests for reprints should be sent to Demetri Psaltis, Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

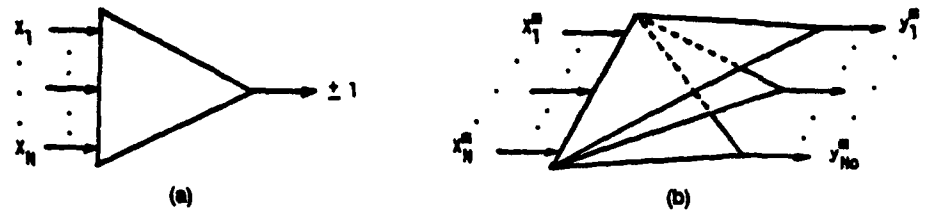


FIGURE 1. (a) Discriminant function; (b) Associative memory constructed as an array of discriminant functions.

the parameters of the memory gives a different mapping. Such a memory cannot learn from the training set to respond in some desirable way to inputs that it has never seen before. The only way to get generalization when $C = D \log_2 K/N_0$ is to impose on it the overall structure of the memory before learning begins. One of the appealing features of neural architectures is the considerable redundancy in the degrees of freedom that is typically available. Therefore, there is hope that while a memory learns specific input–output correspondences it can also discover the underlying structure that may exist in the problem and learn to respond correctly for a set of inputs much larger than the training set. Moreover, the same redundancy is responsible for the error tolerance that is evident in many neural architectures. Higher order memories are generally redundant and they can provide us with a methodology for selecting the degree of redundancy along with the number of degrees of freedom and the associated capacity to store random problems.

It is important to keep in mind that (1) holds for *arbitrary* mappings. If the input and output vectors are restricted in some way that happens to be matched to the architecture of a particular associative memory then it may be possible to overcome this limit. However, selecting the architecture of the associative memory such that it optimally implements only a subset of all possible associations is basically equivalent to choosing the architecture so that it generalizes in a desirable way. For instance suppose that we design an associative memory so that it is shift invariant (i.e., the output is insensitive to a change in the position of the input) (Maxwell *et al.*, 1986; Psaltis & Hong, 1987). Then this system will respond predictably to all the shifted versions of the patterns that were used to train it. We can equivalently think of this system as having a larger storage capacity than the limit of (1) over the set of shift invariant mappings. If we can identify a priori the types of generalization we wish the memory to exhibit, and we can find ways to impose these on the architecture, then this is certainly a sensible thing to do. Higher order memories can also provide a convenient framework within which this can be accomplished.

The penalty we must pay for the increase in the storage capacity that is afforded by the increase in the degrees of freedom in a higher order associative memory is increase in implementation complexity. The computer that implements a higher order memory must have sufficient storage capacity to store a very large number of parameters. Moreover it must be capable of addressing the stored information with a high degree of parallelism in order to produce an output quickly. We will discuss in this paper optical implementations of second order memories and we will show a remarkable compatibility between the computational requirements of these memories and the ability of optics to store information in three dimensions.

1.2 Linear Discriminant Functions and Associative Memories

We will consider as a precursor the most familiar associative memories that are constructed as arrays of linear discriminant functions (Kohonen, 1984). A linear discriminant function is a mapping from the sample space X, a subset of R^N, to 1 or -1.

$$y = \operatorname{sgn}\{\mathbf{w}^t \cdot \mathbf{x} + w_0\}$$

$$= \operatorname{sgn}\{w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_N x_N\} \quad (2)$$

where sgn is the signum function, $\mathbf{w}$ is a weighting vector and w_0 is a threshold value. In this case the capacity is upperbounded by $(N + 1)\log_2 K$ according to our definition of capacity. In this relatively simple case the exact capacity is known to be equal to $C = N + 1$ assuming the input points are in general position and $K = \infty$ (Cover, 1965). An associative memory is constructed by simply forming an array of linear discriminant functions each mapping the same input to a different binary variable. Several algorithms exist for training such memories including the perceptron, Widrow-Hoff, sum of outer products, pseudoinverse, and simplex methods (Duda & Hart, 1973; Hopfield, 1982; Kohonen, 1984; Venkatesh & Psaltis, in press). This memory can be thought of as the first order of the broader class of higher order memories that contain not only a linear expansion of the input vector but also quadratic and higher order terms. We will see in Section 3 that the learning methods that are applicable to the linear memories generalize directly to the higher order memories. First, however, we will describe the properties of the mappings that are implementable with higher order memories in Section 2. Finally, in Section 4 we will describe optical implementations of quadratic optical memories (Psaltis, Park, & Hong, 1986).

2. PROPERTIES OF HIGHER ORDER MEMORIES

A Φ-function is defined to be a *fixed* mapping of the input vector x to an L-dimensional vector z followed by a linear discriminant function.

$$y = \text{sgn}\{\mathbf{w}' \cdot \mathbf{z}(\mathbf{x}) + w_0\}$$

$$= \text{sgn}\{w'_1 z_1 + w'_2 z_2 + \cdots + w'_L z_L + w_0\} \quad (3)$$

where $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots, z_L(\mathbf{x}))$, $\mathbf{w}'$ is an L dimensional weighting vector and $\mathbf{z}(\mathbf{x})$ is an L dimensional vector derived from x. The storage capacity in this case is equal to the capacity of the second layer $L + 1$ (Cover, 1965) if the samples z are in general position whereas the upper bound on the capacity from (1) is $(L + 1)\log_2 K$. The inefficiency in this case is $\log_2 K$ bits, the same as for the linear discriminant function even though the capacity can be raised arbitrarily by increasing L. It is not known what the exact relationship between L and K is, that is, we do not know whether for higher dimensions we need better resolution for the values of the weights to be capable of implementing a fixed fraction of the linear mappings. Recently, Mok and Psaltis (personal communication) have found the asymptotic (large N) statistical capacity to be $C = N$ for a linear discriminant function with binary weights. This result implies that even for large N, for the vast majority of linear dichotomies, a large number of levels is not required. Therefore a Φ-function is an effective and straightforward method for increasing the capacity of an associative memory without loss in efficiency.

A higher order associative memory is an array of Φ-functions with the mappings $\mathbf{z}(\mathbf{x})$ being polynomial expansions of the vector x. The schematic diagram of a higher order associative memory is shown in Figure 2. When the polynomial expansion is of the rth order in x then the output vector y is given by

$$y_l = \text{sgn}\{W'_l(\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}) + W_l^{r-1}(\mathbf{x}, \ldots, \mathbf{x}) + \cdots + W_l^2(\mathbf{x}, \mathbf{x}) + W_l^1\mathbf{x} + w_{l0}\} \quad (4)$$

where $l = 1, \ldots, N_0$, W_l^k is a k-linear symmetric mapping and W_l^1 is equivalent to $\mathbf{w}^l$ in (2). According to (3)

$$z_j(\mathbf{x}) = x_{p_1(j)}^{n_1} x_{p_2(j)}^{n_2} \cdots x_{p_r(j)}^{n_r} \quad (5)$$

where $j = 1, 2, \ldots, L$, $p_i(j) \in \{1, 2, \ldots, N\}$, such that all the $_j$ are distinct, and $n_1, n_2, \ldots, n_r = 0, 1$. Then L is $\binom{N+r}{r}$ (Cover, 1965), and hence the capacity bound is $(\binom{N+r}{r} + 1)\log_2 K$ as before. For example, if $r = 2$, the function becomes quadratic and has the form $y_l = \mathbf{x}^t W_l^2 \mathbf{x} + W_l^1 \mathbf{x} + w_{l0}$ and the number of non-redundant terms in the quadratic expansion is $L = (N + 1)(N + 2)/2$.

The components of the vector z are binary-valued if x is binary. In this case, the samples cannot be assumed to be in general position since there are at most $N + 2$ binary vectors in N dimensional space which lie in general position. We will evaluate the effectiveness of higher order mappings in producing representations $\mathbf{z}(\mathbf{x})$ that are separable by the second layer of weights by calculating the Hamming distance between z vectors given the Hamming distance between the corresponding x vectors. We expect that if the Hamming distance between two binary vectors is large then they are easy to distinguish from one another.

2.1 Complete Polynomial Expansion of Binary Vectors

There are at most 2^N non-redundant terms in any polynomial expansion (4) of a binary vector x in N dimensions. First, we will consider the following Nth order expansion (or equivalently bit production) for the bipolar vectors x in N dimensional binary space $\{1, -1\}^N$:

$$\mathbf{z} = \mathbf{z}(\mathbf{x})$$

$$= (1, x_1, x_2, \ldots, x_N, x_1x_2, \ldots, x_1x_2 \cdots x_N)^t. \quad (6)$$

If we apply a linear discriminant function to the new vectors z, then the capacity becomes 2^N which is equal to the total number of possible input vectors (Psaltis & Park, 1986). In other words this memory is capable of performing *any* mapping of N binary variables to any binary output vector y. Of course the number of weights that are needed to implement this memory grows to 2^N times N_0, the number of bits at the output. In what follows we show that in this extreme case the vectors z become orthogonal to each other.

Theorem 2. If we expand binary vectors $\mathbf{x}^m$ ($m = 1, 2, \ldots, 2^N$) in $X_B = \{1, -1\}^N$ to 2^N dimensional binary

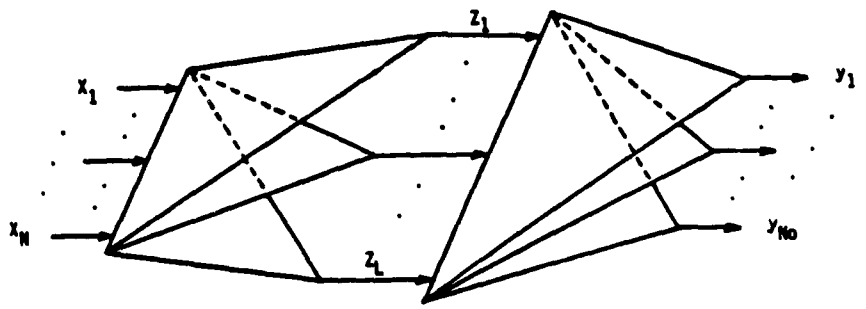


FIGURE 2. Higher order associative memory.

vectors z^m according to (6), where N is the dimensionality of the original feature vectors, then (a) $\langle z^{m_1}, z^{m_2} \rangle = 2^N \delta_{m_1 m_2}$ where $\langle \cdot, \cdot \rangle$ is an inner product, (b) $\sum_i z_i^m = 0$, (c) $\sum_m z_{j_1}^m z_{j_2}^m = 2^N \delta_{j_1 j_2}$ and $\sum_m z_j^m = 0$.

Example: Table 1 is for the case of $N = 3$. Note the orthogonality and the numbers of 1s and -1s in the new vectors and the set of each component of them except the first vector and the set of the first components.

Proof: (a) Let us consider any two different binary vectors in the binary space of $\{1, -1\}^N$ whose Hamming distance is n ($1 \le n \le N$). When they are expanded to two 2^N dimensional binary vectors, the number of kth order terms that have opposite signs in the two expansions is

$$\binom{n}{1}\binom{N-n}{k-1} + \binom{n}{3}\binom{N-n}{k-3} + \binom{n}{5}\binom{N-n}{k-5} + \cdots. \quad (7)$$

Notice that two polynomials have different values if, and only if, they have an odd number of terms whose signs are opposite. The Hamming distance between the two fully expanded (up to order 2^N) vectors can be calculated by adding the number of terms that have different signs over all the orders of the expansion:

$$\begin{aligned}
&\binom{n}{1}\binom{N-n}{0} + \binom{n}{1}\binom{N-n}{1} \\
&\quad + \left\{\binom{n}{1}\binom{N-n}{2} + \binom{n}{3}\binom{N-n}{0}\right\} \\
&\quad + \left\{\binom{n}{1}\binom{N-n}{3} + \binom{n}{3}\binom{N-n}{1}\right\} + \cdots \\
&\quad + \left\{\binom{n}{1}\binom{N-n}{N-n} + \binom{n}{3}\binom{N-n}{N-n-2}\right. \\
&\quad \left. + \binom{n}{5}\binom{N-n}{N-n-4} + \cdots\right\} \\
&\quad + \left\{\binom{n}{3}\binom{N-n}{N-n-1} + \binom{n}{5}\binom{N-n}{N-n-3}\right. \\
&\quad \left. + \binom{n}{7}\binom{N-n}{N-n-5} + \cdots\right\} \\
&\quad + \cdots \\
&= \sum_{i=\text{odd}} \binom{n}{i} \sum_{j=0}^{N-n} \binom{N-n}{j} = 2^{n-1} 2^{N-n} \\
&= 2^{N-1}. \qquad (8)
\end{aligned}$$

The fact that the Hamming distance is 2^{N-1} for any two expanded vectors (for any n) proves that all of the 2^N vectors become orthogonal and that $\langle z^{m_1}, z^{m_2} \rangle = 2^N \delta_{m_1 m_2}$. (b) Just think of the cases where one of the two vectors is $(1, 1, \ldots, 1)$. Then, all the other vectors z have equal number of 1s and -1s because their Hamming distances are all 2^{N-1} from the $(1, 1, \ldots, 1)$ vector. (c) See Duda and Hart (1973, p. 109).

Slepian has discussed this orthogonalization property as a method for designing orthogonal codes and has given a different proof for it (Slepian, 1956). The proof presented here is useful for characterizing higher order memories because it allows us to trace the contribution of each order of the expansion to the orthogonalization and immediately derive results about the properties of quadratic and cubic memories. The output vector y is

$$y_l = \operatorname{sgn}\{W_l \cdot z\} = \operatorname{sgn}\left\{\sum_{i=1}^{2^N} W_{li} z_i\right\} \quad (9)$$

where $l = 1, \ldots, N_0$ and W_l is a 2^N dimensional weighting row vector. The matrix W_{li} that can implement the $x^m \mapsto y^m$ mapping for $m = 1$ to 2^N can be formed in this case simply as the sum of outer products of y^m and z^m:

$$W_{li} = \sum_{m=1}^{2^N} y_l^m z_i^m. \quad (10)$$

2.2 Expansions of a Single Order

The orthogonalization property of the full expansion is interesting because it shows that higher order memories provide a complete framework that takes us from the simplest "neuron," the linear discriminant function, to the full capability of a Boolean look-up table. Higher order memories can indeed provide a valuable tool for designing digital programmable logic arrays. In this paper, however, we are interested in associative memories that are capable of accepting inputs with

TABLE 1

x_1	x_2	x_3	1	x_1	x_2	x_3	x_1x_2	x_2x_3	x_3x_1	$x_1x_2x_3$
1	1	1	1	1	1	1	1	1	1	1
1	1	−1	1	1	1	−1	1	−1	−1	−1
1	−1	1	1	1	−1	1	−1	−1	1	−1
1	−1	−1	1	1	−1	−1	−1	1	−1	1
−1	1	1	1	−1	1	1	−1	1	−1	−1
−1	1	−1	1	−1	1	−1	−1	−1	1	1
−1	−1	1	1	−1	−1	1	1	−1	−1	1
−1	−1	−1	1	−1	−1	−1	1	1	1	−1

large N (e.g., if $N = 10^3$ then $2^N \approx 10^{300}$) in which case considering a full expansion of the input data is completely out of the question. In such cases we are really interested in an expansion that contains a large enough number of terms to provide the capacity needed to learn the problem at hand. In this subsection we analyze the properties of partial expansions that include all the terms of one order.

We will first consider the memory consisting of all the terms of a quadratic expansion with binary input vectors.

$$y_l = \operatorname{sgn}\{\sum_i \sum_j w_{lij} x_i x_j\}$$

$$= \operatorname{sgn}\{\sum_{k=1}^{L} w'_{lk} z_k\}. \tag{11}$$

The number of non-redundant terms in a quadratic expansion of a binary vector is $L = N(N-1)/2$. Let two input vectors have a Hamming distance n. The angle between these two vectors is given by the relation $\cos\theta_1 = 1 - (2n/N)$. The angle θ_2 between the corresponding $z(x)$ vectors can be readily calculated since we know their Hamming distance from the proof of Theorem 2(a):

$$\cos\theta_2 = 1 - \frac{4n(N-n)}{N(N-1)}$$

$$\approx 1 - 4\rho + 4\rho^2 = (1-2\rho)^2 \tag{12}$$

where $\rho = n/N$. θ_2 and θ_1 are plotted versus ρ in Figure 3a. For $\rho < .5$, θ_2 is always larger than θ_1. Specifically for $\rho \ll 1$, $\theta_2 = \sqrt{2} \times \theta_1$. We see therefore that the quadratic mapping not only expands the dimensionality which provides capacity but also spreads the input samples apart, a generally desirable property. For $\rho > .5$ the quadratically expanded vectors are closer to each other than the original vectors and in the extreme case $n = N$, θ_2 becomes zero. This insensitivity of the quadratic mapping to a change in sign of all the bits is a property that is shared by all even order expansions. Next we consider a cubic memory

$$y_l = \operatorname{sgn}\{\sum_i \sum_j \sum_k w_{lijk} x_i x_j x_k\}$$

$$= \operatorname{sgn}\{\sum_{n=1}^{L} w'_{ln} z_n\} \tag{13}$$

where $L = \binom{N}{3} + N$. In Figure 3b we plot θ_3, the angle between two cubically expanded binary vectors as a function of ρ. For convenience, θ_1 is also plotted in the same figure. In this case θ_3 increases faster with ρ for $\rho < .5$. For $\rho \ll 1$, $\theta_3 = \sqrt{3} \times \theta_1$. At $\rho \approx .4$ the cubic expansion gives essentially perfectly orthogonal vectors while for $\rho > .5$, θ_3 remains smaller than θ_1 and in the limit $\rho = 1$, $\theta_3 = \pi$. Thus the cubic memory discriminates between a vector and its complement.

The basic trends that are evident in the quadratic and cubic memories generalize to any order r. The number of independent terms in the rth order expansion of a binary vector is $\binom{N}{r}$ which is maximum for $r \approx N/2$. Again this is not of practical importance because the number of terms in a full expansion of this sort is prohibitively large. What is of interest however is the effectiveness with which relatively small order expansions can orthogonalize a set of input vectors. The angle θ_r between two vectors that have been expanded to the rth order is given by the following relation:

$$\cos\theta_r = \frac{\binom{N}{r} - 2\sum_{i=\text{odd}} \binom{n}{i}\binom{N-n}{r-i}}{\binom{N}{r}}. \tag{14}$$

We can obtain a simpler expression for the interesting case $r \ll N$ and for small ρ, $\theta_r \approx \sqrt{r} \times \theta_1$.

Proposition 3: *For* $r \ll N$,

$$\cos\theta_r \approx (1-2\rho)^r. \tag{15}$$

Moreover, for small ρ,

$$\theta_r \approx \sqrt{r}\theta_1 \tag{16}$$

where $\theta_1 \approx 2\sqrt{\rho}$.

Proof: For a small r, we can make the approximations $\binom{N}{r} \approx N^r/r!$, $\binom{n}{i} \approx n^i/i!$, and $\binom{N-n}{r-i} \approx (N-n)^{r-i}/(r-i)!$. Then, $\cos\theta_r$ is approximated as follows:

$$\cos\theta_r \approx 1 - 2\sum_{i=\text{odd}} \frac{r!}{i!(r-i)!}\rho^i(1-\rho)^{r-i}$$

$$= (1-2\rho)^r$$

because of these relationships:

$$\sum_{i=\text{odd}} + \sum_{i=\text{even}} = (1-\rho+\rho)^r = 1,$$

$$\sum_{i=\text{odd}} - \sum_{i=\text{even}} = -(1-\rho-\rho)^r = -(1-2\rho)^r.$$

When $\rho \ll 1$, $\cos\theta_r$, which is approximately $1 - \theta_r^2/2!$, is approximated by $1 - 2r\rho$ directly from (14) or from (15). Therefore, it is followed by (16) that $\theta_r \approx 2\sqrt{r\rho}$.

We plot θ_r versus ρ for selected orders in Figure 4 using (15). It is evident that increasing r results in better separated feature vectors. Polynomial mappings act as an effective mechanism for increasing the dimensionality of the space in which inputs are classified because they guarantee a very even distribution of the samples in this new space.

3. TRAINING OF HIGHER ORDER MEMORIES

Once the initial polynomial mapping has been selected, the rest of the system in a higher order memory is simply a linear discriminant function. As such it can be trained by any of the existing methods for training

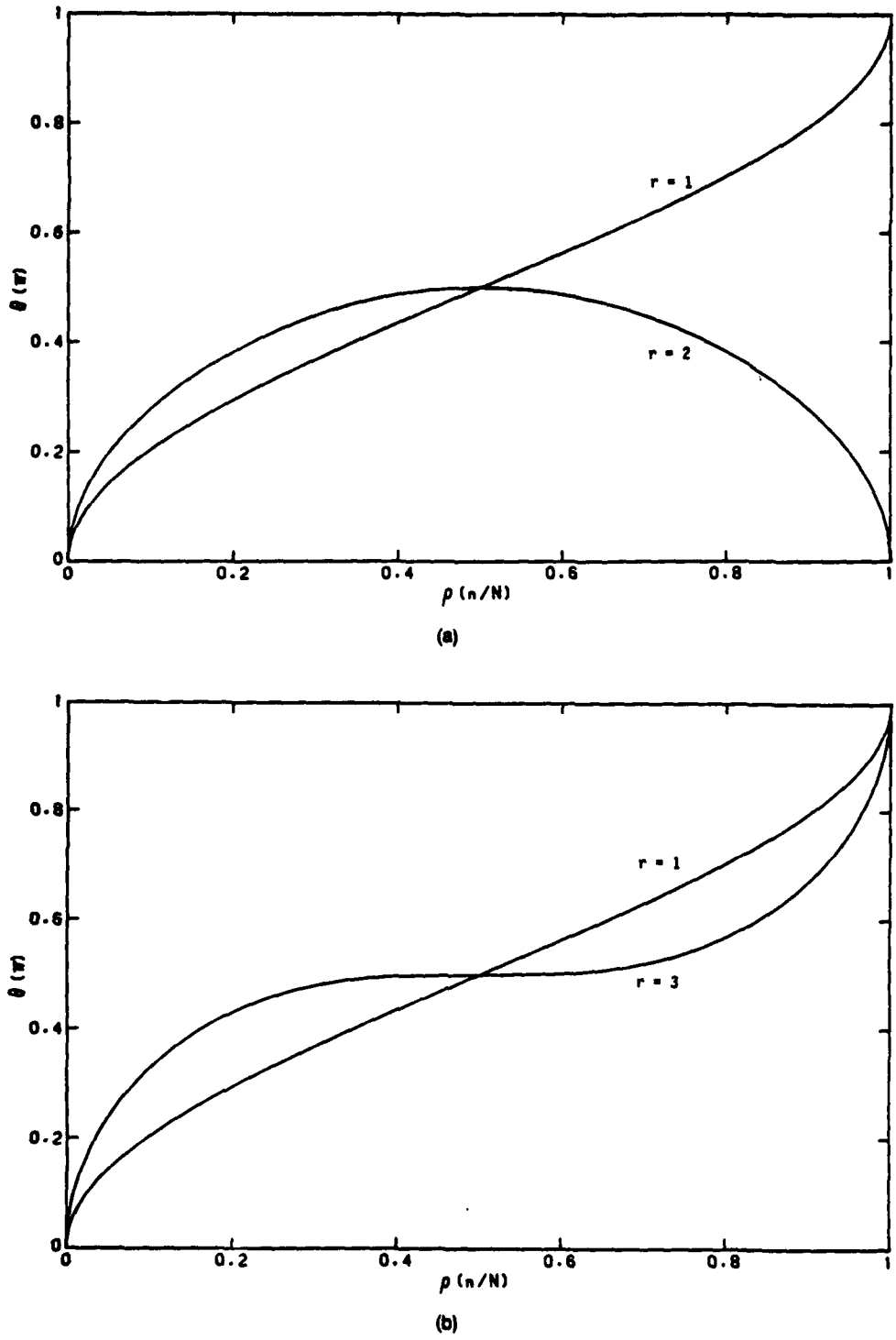


FIGURE 3. (a) The angle between linearly and quadratically expanded vectors as a function of the hamming distance at the input; (b) The angle between linearly and cubically expanded vectors as a function of the hamming distance at the input.

linear discriminant functions. For instance the pseudoinverse (Kohonen, 1984; Venkatesh & Psaltis, in press) can be used to calculate the set of weights that will map a set of L-dimensional expanded vectors $\mathbf{z}^m$ to the associated output vectors $\mathbf{y}^m$. Alternately, error driven algorithms such as the perceptron or adaline can be used to iteratively train the memory by repeatedly presenting the input vectors to the system, monitoring the output to obtain an error signal, and modifying the weights so as to gradually decrease the error.

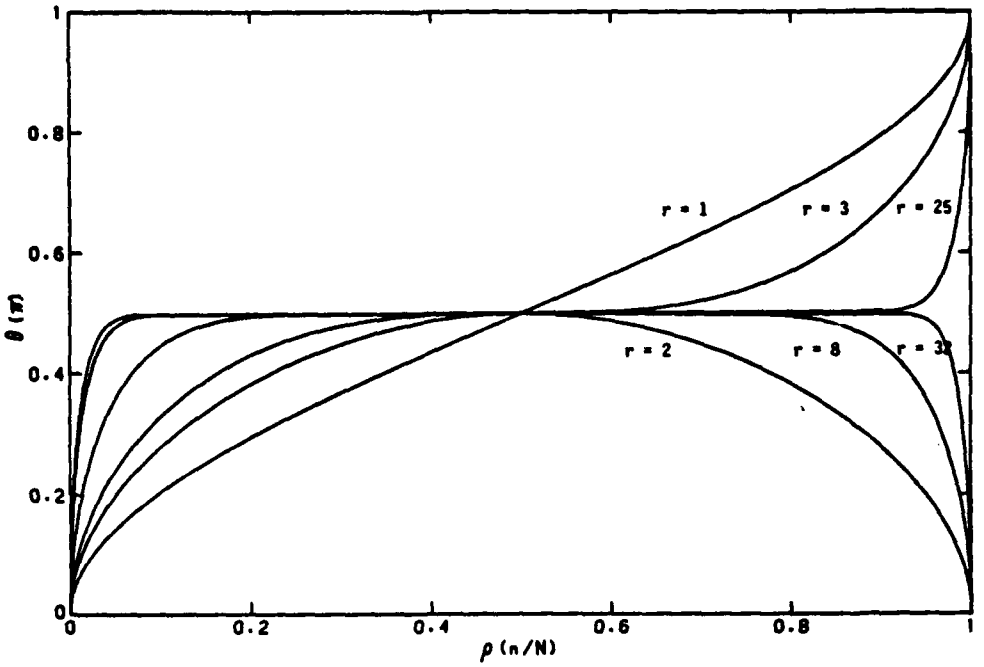


FIGURE 4. The angle between expanded vectors for selected orders.

The relative ease with which higher order memories can be trained is a very important advantageous feature of this approach. A higher order memory is basically a multilayered network where the first layer is selected a priori. In terms of capacity alone, there is no advantage whatsoever in having multiple layers with modifiable weights. From Theorem 1 we know that at best the capacity is determined by the number of modifiable weights. For a higher order memory we get the full advantage of the available degrees of freedom whereas if we put the same number of weights in multiple layers the resulting degeneracies will decrease the capacity. The relative advantage of trainable multiple layers is the potential for generalization that emerges through the learning process. The generalization properties of higher order memories on the other hand are mostly determined by the choice of the terms used in the polynomial expansion in the fixed first layer. Thus the generalization properties of these memories as described in this paper are imposed a priori by the designer of the system.

The sum of outer products algorithm that has been used extensively for training linear associative memories can also be used for training the higher order memories and this algorithm generalizes to the higher order case in particularly interesting ways. In addition, this particular learning algorithm is predominantly used for the holographic optical implementations that are described in the following section. Therefore we will discuss in some detail the properties of higher order memories that are trained using this rule.

3.1 The Outer Product Rule

Let us consider associative memories constructed as an expansion of the r-order only with input samples in an N dimensional binary space and $r \geq 1$.

$$y_l = \operatorname{sgn}\{\sum_{j_1 j_2 \cdots j_r} W_{l j_1 j_2 \cdots j_r} x_l x_{j_1} x_{j_2} \cdots x_{j_r}\}, \quad (17)$$

where $1 \leq j_1, j_2, \ldots, j_r \leq N$, $1 \leq l \leq N_0$. The number of independent terms L in the rth order expansion is $\binom{N+r-1}{r}$ which for $r \ll N$ can be approximated by $N^r/r!$

The expression for the weights of the rth order expansion using the sum of outer products algorithm is (Chen et al., 1986; Psaltis & Park, 1986)

$$W_{l j_1 j_2 \cdots j_r} = \sum_{m=1}^{M} y_l^m x_{j_1}^m x_{j_2}^m \cdots x_{j_r}^m \quad (18)$$

where M is the number of vectors stored in the memory, $\mathbf{y}^m$ is an output vector associated with an input vector $\mathbf{x}^m$ as before. With the above expression for the weight tensor (17) can be rewritten as follows

$$y_l = \operatorname{sgn}\{\sum_{m=1}^{M} y_l^m (\sum_{j=1}^{N} x_j^m x_j)^r + w_l^0\}. \quad (19)$$

The above equation suggests an alternate implementation for higher order memories that are trained using the outer product rule. This is shown schematically in Figure 5. The inner products between the input vector and all the stored vectors $\mathbf{x}^m$ are formed first, then raised to the rth power, and the signal from the mth unit is

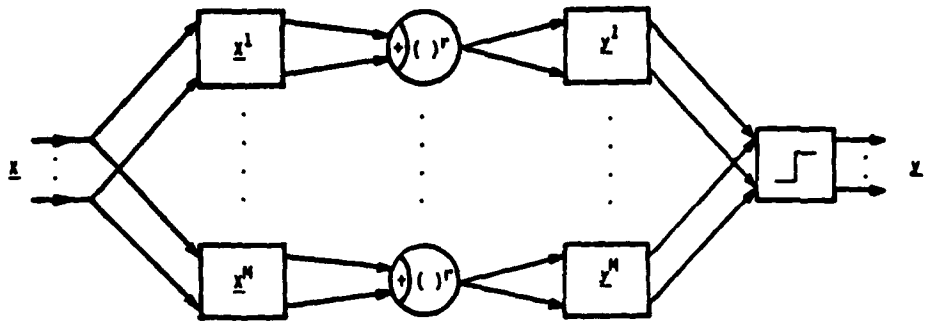


FIGURE 5. Outer product, rth order associative memory.

connected to the output through interconnective weights y_l^m. If $\mathbf{y}^m = \mathbf{x}^m$ then the memory is autoassociative, and in this case the output can be fed back to the input resulting in a system whose stable states are programmed to be the vectors $\mathbf{x}^m$. This becomes a direct extension of the Hopfield network (Anderson, 1983; Hopfield, 1982; Nakano, 1972) to the higher order case. Assuming that $\mathbf{x} = \mathbf{x}^n$ is one of the stored vectors, y_l becomes

$$y_l = \operatorname{sgn}\{N^r y_l^n + \sum_{m \neq n} y_l^m \Big(\sum_{j=1}^{N} x_j^m x_j^n\Big)^r + w_l^0\}$$

$$= \operatorname{sgn}\{N^r y_l^n + n_l(\mathbf{x}^n)\} \tag{20}$$

where the first term is the desired signal term and n_l is a noise term. The threshold weight is set to zero.

The expectation value of $n_l(\mathbf{x}^n)$ is zero if the bits that comprise the stored binary input and output vectors are drawn randomly and independently having equal probability of being $+1$ or -1. If this is the case then

$$E\Big(\sum_{u'} x_l^m x_l^{m'}\Big) = \sum_{u'} \delta_{u'},$$

$$E\Big(\sum_{mm'} x_l^m x_l^{m'}\Big) = \sum_{mm'} \delta_{mm'} \tag{21}$$

where δ_{ij} is the Kronecker delta function. The variance of n_l is calculated as follows:

$$E(n_l^2)$$

$$= E\Big(\sum_{m \neq n} \sum_{m' \neq n} y_l^m y_l^{m'} \sum_{j_1 j_2 \cdots j_r} \sum_{s_1 s_2 \cdots s_r} x_{j_1}^m x_{j_2}^m \cdots$$

$$x_{j_r}^m x_{j_1}^n x_{j_2}^n \cdots x_{j_r}^n x_{s_1}^{m'} x_{s_2}^{m'} \cdots x_{s_r}^{m'} x_{s_1}^n x_{s_2}^n \cdots x_{s_r}^n\Big)$$

$$= E\Big(\sum_{m \neq n} \sum_{j_1 j_2 \cdots j_r} \sum_{s_2 \cdots s_r} x_{j_1}^m x_{j_2}^m \cdots x_{j_r}^m x_{s_1}^m x_{s_2}^m \cdots$$

$$x_{s_r}^m x_{j_1}^n x_{j_2}^n \cdots x_{j_r}^n x_{s_1}^n x_{s_2}^n \cdots x_{s_r}^n\Big). \tag{22}$$

In the above we used the facts that different stored vectors are uncorrelated (i.e., for $m \neq m'$) and $y_l^2 = 1$. Then, the variance becomes $(M-1)Q(N, r)$, where $Q(N, r)$ is the number of possible permutations such that

$$\delta_{i_1 t_1} \delta_{i_2 t_2} \cdots \delta_{i_r t_r} = 1 \tag{23}$$

where the set of variables $\{i_1, \ldots, i_r, t_1, \ldots, t_r\}$ spans all the combinations produced by the set of variables $\{j_1, \ldots, j_r, s_1, \ldots, s_r\}$. The variance can be calculated exactly for the cases $r = 1$, 2, and 3 and it is $(M-1)N$, $(M-1)(3N^2 - 2N)$ and $(M-1)(15N^3 - 30N^2 + 16N)$, respectively. For the general case we will derive lower and upper bounds which for large N provide us with a good estimate of the variance for any order r.

Proposition 4: The total number of permutations, $Q(N, r)$, for which (23) holds, satisfies the relationship:

$$P(N, r)\frac{(2r)!}{2^r r!} + \binom{2r}{4} P(n, r-1)\frac{(2r-4)!}{2^{r-2}(r-2)!}$$

$$\leq Q(N, r) \leq N^r \frac{(2r)!}{2^r r!} \tag{24}$$

where $P(m, n) = m!/(m-n)!$.

Proof: The number of ways of making r pairs of $2r$ items is $(2r-1)(2r-3)\cdots(3)(1) = (2r)!/2^r r!$. The items that we are concerned with are the variables i_j, t_j and each of these variables can take one of N values. We can only select the values of half these variables (N^r possibilities) and for each of these choices we can create r pairs. Hence the upperbound is $N^r(2r)!/2^r r!$. This is an upper bound because we have overcounted for different pairings of variables that have the same value.

The initial lower bound is derived if each pair has a different value from all others, which eliminates the possibility of overcounting. The number of possible ways to satisfy (23) with the variables in any two pairs not taking the same values is $P(N, r)(2r)!/2^r r!$. This is an underestimate because all pairs that contain variables taking the same value should be counted once. We can thus improve the lower bound by counting the number of ways these degenerate pairings occur and adding them into the previous bound. For example when two pairs out of r have the same values with $\binom{2r}{4}$ choices, there are $\binom{2r}{4}NP(N-1, r-2)(2r-4)!/2^{r-2}(r-2)!$ possible permutations where $(2r-4)!/2^{r-2}(r-2)!$ is the number of ways of making $r-2$ pairs of $2r-4$ items. Therefore, $Q(N, r)$ is lower bounded by $P(N, r)(2r)!/2^r r! + \binom{2r}{4}P(N, r-1)(2r-4)!/2^{r-2}(r-2)!$, since $NP(N-1, r-2) = P(N, r-1)$.

We can get a very good approximation to the *SNR* using the approximations of $M - 1 \approx M$ and $Q(N, r) \approx N^r(2r)!/2^r r!$ which are very nearly true for the interesting case $r \ll N$:

$$SNR \approx \frac{N^r}{\{MN^r(2r)!/2^r r!\}^{1/2}} = \left\{\frac{N^r}{M}\frac{2^r r!}{(2r)!}\right\}^{1/2}. \quad (25)$$

For example, the linear memory, $r = 1$, has a *SNR* $\approx (N/M)^{1/2}$, the quadratic memory, $r = 2$, a *SNR* of $N/(3M)^{1/2}$ and the cubic memory, $r = 3$, a *SNR* of $(N^3/15M)^{1/2}$. We can obtain an estimate for the capacity of an rth order memory by equating the signal to noise ratios of the linear and rth order memories and solving for M_r, the number of stored vectors that will yield the equality. For r, small compared to N, we obtain

$$\frac{M_r}{M_1} = N^{r-1}\frac{2^r r!}{(2r)!}. \quad (26)$$

Comparing its value with the capacity M_1 of a linear memory we can obtain the relationship between the capacities, that is, $M_r/M_1 = N^{r-1}2^r r!/(2r)!$. For example M_2 of a quadratic memory is $M_1 N/3$ and M_3 of a cubic memory is $M_1 N^2/15$.

The diagonal terms in a high order memory $W_{ij_1j_2\cdots j_r}$ can be defined as those of which all the indexes j are not different. We form the weight tensor with zero diagonal as follows:

$$W_{ij_1j_2\cdots j_r} = \begin{cases} \sum_m y_i^m x_{j_1}^m x_{j_2}^m \cdots x_{j_r}^m & \text{if } j\text{s are all different,} \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

When the input is one of the stored vectors $\mathbf{x}^n$ and the weight tensor has zero diagonal, the output y_i becomes

$$\begin{aligned} y_i &= \operatorname{sgn}\Big\{\sum_{\text{different } j} W_{ij_1j_2\cdots j_r} x_{j_1}^n x_{j_2}^n \cdots x_{j_r}^n + w_i^0\Big\} \\ &= \operatorname{sgn}\{P(N, r)y_i^n \\ &\quad + \sum_{m\neq n} y_i^m \sum_{\text{different } j} x_{j_1}^m x_{j_2}^m \cdots \\ &\qquad x_{j_r}^m x_{j_1}^n x_{j_2}^n \cdots x_{j_r}^n + w_i^0\} \end{aligned} \quad (28)$$

where the first term is a signal term and the second a noise term as before. The variance of the noise term is easily shown to be $(M - 1)P(N, r)r!$ using (21). Therefore, the *SNR* becomes

$$SNR = \left\{\frac{P(N, r)}{(M-1)r!}\right\}^{1/2} \approx \left\{\binom{N}{r}\Big/M\right\}^{1/2} \quad (29)$$

which can be approximated as $(N^r/Mr!)^{1/2}$ for $r \ll N$.

Chen and his coworkers (1986) introduced an energy function (Cohen & Grossberg, 1983; Hopfield, 1982) for the rth order autoassociative memory with feedback and outer products as follows:

$$E_r = -\sum_{m=1}^{M} \langle \mathbf{x}^m, \mathbf{x}\rangle^{r+1} \quad (30)$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product of two vectors. The change in the energy due to a change $\delta\mathbf{x}$ in the state of the network was shown by Chen *et al.* (1986) to be decreasing for odd r.

$$\begin{aligned} \Delta E_r &= E_r(\mathbf{x} + \delta\mathbf{x}) - E_r(\mathbf{x}) \\ &= -(r+1)\sum_i \delta x_i \sum_{j_1\cdots j_r} W_{ij_1j_2\cdots j_r} \\ &\qquad \times x_{j_1}x_{j_2}\cdots x_{j_r} - R_r \end{aligned} \quad (31)$$

where

$$R_r = \sum_m \sum_{j=2}^{r+1} \binom{r+1}{j} \langle \mathbf{x}^m, \mathbf{x}\rangle^{r+1-j}\langle \mathbf{x}^m, \delta\mathbf{x}\rangle^j. \quad (32)$$

The first term in (31) is always nonpositive because of the specification of the update rule: $\delta x_i \geq 0$ if $\sum_{j_1\cdots j_r} \times W_{ij_1j_2\cdots j_r}x_{j_1}x_{j_2}\cdots x_{j_r} \geq 0$ and vice versa. Chen *et al.* (1986) showed that the second term is also nonpositive by showing that R_r is an increasing function of r for r odd and $R_1 > 0$.

For r even it is possible to prove the autoassociative memory converges only for asynchronous updating even though in simulations even order autoassociative memories consistently converge as well. The fact that the energy is not always decreasing when r is even may actually be helpful for getting out of local minima and settling in the programmed stable state which are global minima in a region of the energy surface. A descent procedure that is always decreasing in energy cannot escape local minima since there is no mechanism for climbing out of them. As an example, consider a quadratic memory, that is, $r = 2$ (even), whose energy function is given by

$$E_2 = -\sum_{ijk} W_{ijk}x_i x_j x_k \quad (33)$$

$$\begin{aligned} \Delta E_2 &= -3\sum_{ijk} W_{ijk}x_j x_k \delta x_i - 3\sum_{ijk} W_{ijk}x_k \delta x_i \delta x_j \\ &\qquad - \sum_{ijk} W_{ijk}\delta x_i \delta x_j \delta x_k. \end{aligned} \quad (34)$$

The first term is nonincreasing but the second and third terms can be increasing. If the vector $\mathbf{x}$ is very close to one of the stored vectors $\mathbf{x}^n$ then the first term becomes dominant and the energy will be very likely to be nonincreasing causing the system to settle at $\mathbf{x} = \mathbf{x}^n$. If $\mathbf{x}$ is not close to any of the stored vectors, then all three terms in the above equations are on the average comparable to each other and since two of them are not nondecreasing the energy function may be increasing and it is possible to escape from local minima.

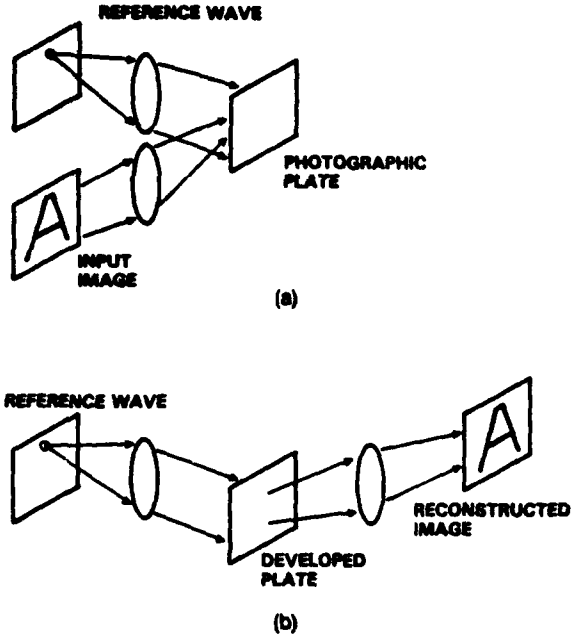


FIGURE 6. Holographic recording and reconstruction. (a) Recording, (b) reconstruction.

4. OPTICAL IMPLEMENTATIONS OF QUADRATIC ASSOCIATIVE MEMORIES

The outer product quadratic associative memories described in the previous section require three basic components for their implementation: interconnective weights, a square-law device, and a threshold nonlinearity. In this section, we present a variety of optical implementations using either planar or volume holograms to provide the interconnection pathways and optical or electro-optical devices to provide the required nonlinearities.

Since holographic techniques are used to implement the required interconnections, we will first briefly discuss holography (Collier, Burkhardt, & Lin, 1971) and in particular the distinction between the use of planar versus volume holograms. The holographic process is shown schematically in Figure 6. In the recording step (Figure 6a) the interference between the reference plane wave that is created by collimating the light from a point source using a lens and the wave originating from the object "A" is recorded on a planar light sensitive medium such as a photographic plate. When the developed plate is illuminated with the same reference wave, the field that is diffracted by the recorded interference pattern gives a virtual image of the original object which can be converted to a real image with a lens. The reconstruction of the hologram is thus equivalent to interconnecting the single point from which the plane wave reference is derived to all the points that comprise the reconstructed image. The weight of each interconnection is specified by the interference pattern stored in the hologram.

Volume holograms are prepared and used in the same manner except that whereas a planar hologram records the interference pattern as a two dimensional pattern on a plane, a volume hologram records the interference pattern throughout the volume of a three dimensional medium. The disparity in the dimensionalities of the two storage formats results in marked differences in the capabilities of the two processes. This difference is explained with the aid of Figures 7a and

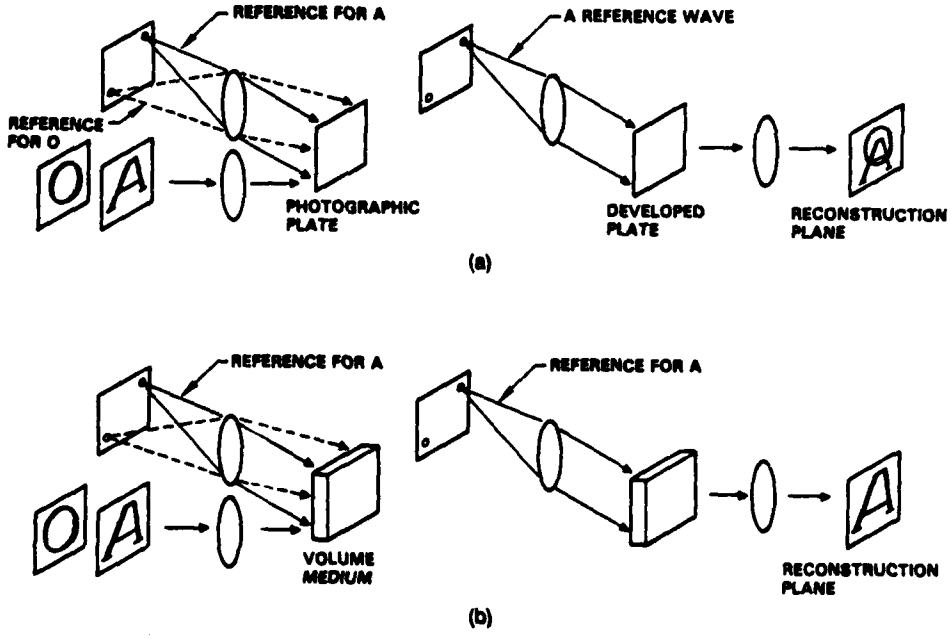


FIGURE 7. Holographic interconnections using (a) planar versus (b) volume holograms.

7b where the reconstruction of both a planar and a volume hologram are shown. Each hologram is prepared to store the two images "A" and "O" by double exposure with each image being associated with a reference plane wave that is incident on the hologram at a different angle. Each reference plane wave is generated by a separate point source and thus the reconstruction of a hologram with the two reference waves is equivalent to interconnecting multiple input points to all the points on the plane of the reconstructed image. In the case of the planar hologram, however, when either one of the reference waves is incident both images are reconstructed. This implies that we cannot in this case independently specify how each of the input points is connected to the output. In contrast, because of the interaction of the fields in the third dimension (Kogelnik, 1969) the volume hologram is able to resolve the differences in the angle of incidence of the reference beam and upon reconstruction when the reference for "A" illuminates the medium, only "A" is reconstructed and similarly for the second pattern. When both input points are on simultaneously then each is interconnected to the output independently according to the way it was specified by the recording of the two holograms. Thus volume holograms provide more flexibility for implementing arbitrary interconnections which translates to efficient three dimensional storage of the interconnective weights needed to specify the quadratic memory.

Another way in which we can draw the distinction between planar and volume holograms is in terms of the degrees of freedom. The implementation of a quadratic memory whose input word size is N bits requires approximately N^3 interconnections for the three dimensional interconnection tensor. The number of degrees of freedom of the planar hologram of area A is upper bounded by A/δ^2 while that of a volume hologram is limited to V/δ^3, where V is the volume of the crystal and δ is the minimum detail that can be recorded in any one dimension (Psaltis, Yu, Gu, & Lee, 1987; Van Heerden, 1963). Equating the degrees of freedom that are required to do the job to those that are available, the crystal volume is determined to be at least $V = N^3\delta^3$ whereas a planar hologram to do the same job would require a hologram of area $A = N^3\delta^2$. For comparison, a network with $N = 10^3$ can in principle be implemented using a cubic crystal with the length of each side being $l_v = N\delta = 1$ cm, but a square planar hologram is required to have the length of each side be at least $l_p = N^{3/2}\delta = 0.33$ m at $\delta = 10$ μm. Thus, the volume hologram offers a more compact means of implementing large memory systems.

4.1 Volume Hologram Systems

There are several schemes for fully utilizing the interconnective capability of volume holograms (Psaltis

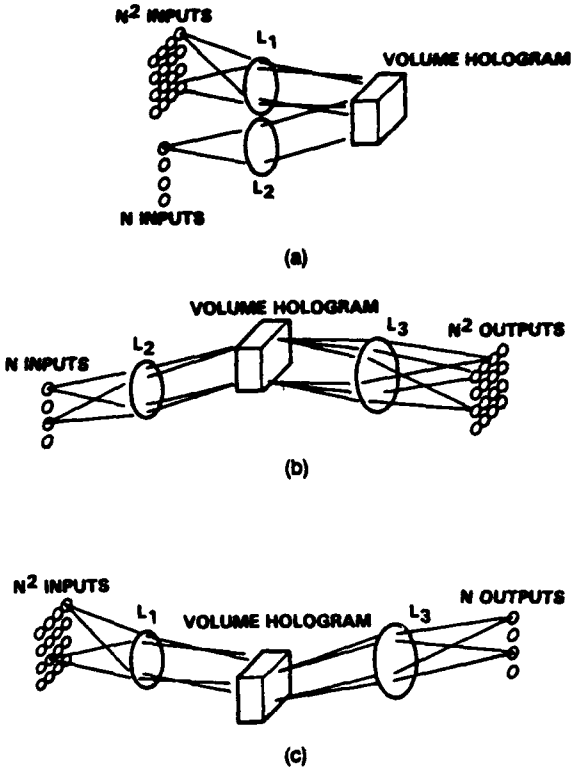


FIGURE 8. Optical interconnections using volume holograms. (a) Recording apparatus; (b) $N \mapsto N^2$ mapping; (c) $N^2 \mapsto N$ mapping.

et al., 1987; Psaltis, Brady, & Wagner, in press). For the implementation of quadratic memories we use volume holograms to fully interconnect a 2-D pattern to a 1-D pattern ($N^2 \mapsto N$ mappings) and also the reverse ($N \mapsto N^2$). The geometry for recording the weights for both cases is shown in Figure 8a and the reconstruction geometries are illustrated in Figures 8b and 8c. The circles represent the resolvable spots at the various planes in the system. The waves emanating from each point at the input planes are transformed into plane waves by the Fourier transform lenses L_1 and L_2 and interfere within the crystal, creating volume gratings.

The weights are loaded into the volume hologram with multiple holographic exposures in the system of Figure 8a. In the following subsections we will describe several specific procedures for doing so. For the $N \mapsto N^2$ mapping (Figure 8b) in reading out the stored information, a single source in the input array reconstructs one of the N 2-D images consisting of N^2 pixels that it is associated with. The rest of the images, which belong to the other input points, are not read out because of the angular discrimination of volume holograms. The counterpart to this scheme, shown in Figure 8c, implements an arbitrary $N^2 \mapsto N$ mapping. This setup is basically the same as that of Figure 8b except that the roles of the input planes have been interchanged or equivalently the direction in which light propagates has been reversed.

4.1.1 $N^2 \mapsto N$ *Schemes.* First, we consider a method by which the full three dimensional interconnection tensor is implemented directly with a volume hologram. Recall that if the weight tensor is trained using the sum of outer products then it is given by

$$w_{ijk} = \sum_{m=1}^{M} y_i^m x_j^m x_k^m, \tag{35}$$

where x_i^m represents the mth input memory vector and y_i^m represents the associated output vector. Such a memory is accessed by first creating an outer product of the input vector and multiplying it with w_{ijk} as follows:

$$y_i = \text{sgn}\{\sum_{j=1}^{N}\sum_{k=1}^{N} w_{ijk} x_j x_k\}. \tag{36}$$

The volume hologram is prepared using the setup in Figure 8a. First, the outer product matrix of the mth memory input vector, $x_j^m x_k^m$, is formed on an electronically addressed spatial light modulator (SLM) (Warde & Fisher, 1987). Another one-dimensional SLM whose transmittance represents the mth output vector y_i^m is placed in the other input plane, and the two SLMs are illuminated by coherent light. The transmitted waves are then Fourier transformed by lenses L_1 and L_2 to interfere within the crystal volume to create index gratings. This procedure is repeated for all M associated input–output pairs so that a sum of M holograms is created in the crystal. For the quadratic outer product memory whose capacity is fully expended, this involves on the order of $N^2/\log N$ exposures.

We will now describe another method for recording the weight vector in the volume hologram that involves fewer exposures and can also be used not only for the outer product scheme but for recording any given weight tensor as well. The same basic recording architecture of Figure 8a is used in this case also. In the first exposure, the top light source in the linear array is turned on while the SLM is programmed with the matrix w_{1jk}, where w_{ijk} is the interconnection tensor. When the SLM is illuminated with light coherent with that of the point source, the crystal records the mutual interference pattern as a hologram of the image w_{1jk} with a reference beam that is the plane wave generated from the top light source. In the next step, the second source is turned on while the SLM is programmed with the matrix w_{2jk}. In this manner the connectivity for all the points in the linear array at the input are sequentially specified and the memory training is completed when all N exposures have been made. The disadvantage of this method relative to the outer product recording is the need to precalculate electronically the weight tensor but it has the advantage of fewer exposures (N versus $N^2/\log N$) and greater flexibility in choosing the training method.

The architecture in Figure 8c is used to access the data stored in the hologram by either one of the recording methods described above. The electronically addressed 2-D SLM is placed at the input plane and it is programmed with the outer product matrix $x_k x_j$ of the input vector. The light from the N^2 input points is interconnected with the N output points via the recorded w_{ijk} interconnect kernel. A linear array of N photodetectors is positioned to sample the output points.

It is important to restate at this juncture that this particular implementation achieves the quadratic interconnections by first transforming the N input features (i.e., the N elements of the input vector x_j) into a set of N^2 features via the outer product operation. The result is that although the interconnections are quadratic with respect to the N original feature points, they are linear with respect to the N^2 transformed features. This allows the application of error driven learning algorithms for linear networks such as the *Adaline* (Widrow & Hoff, 1960) where the interconnections are developed by an iterative training process. The operation of such a learning scheme is illustrated in Figure 9 which is the same basic architecture as Figure 8c with feedback from the output back into one of the input ports. Each iteration consists of a reading and a writing phase. During the reading phase, the interconnections present in the crystal are interrogated with a particular item to be memorized by illuminating the 2-D SLM which contains the outer product matrix $\mathbf{x}^m\mathbf{x}^m$ and the output is formed on the detector array. In the subsequent writing phase, the error pattern generated by subtracting the actual output from the desired output pattern is loaded into the 1-D SLM and both SLMs (the 2-D SLM still contains $\mathbf{x}^m\mathbf{x}^m$) are illuminated with coherent light, forming a set of gratings in addition to the previously recorded gratings. The procedure is iteratively repeated for each item to be memorized until the output error is sufficiently small. This algorithm is a descent procedure designed to minimize the mean squared cost $\epsilon = \frac{1}{M}\sum_{m=1}^{M}[\sum_{j=1}^{N}\sum_{k=1}^{N} w_{ijk}x_j^m x_k^m - y^m]^2$ by iteratively updating the interconnection values.

4.1.2 $N \mapsto N^2$ *Schemes.* The $N \mapsto N^2$ mapping capability of the volume hologram which is the inverse of that required for the architectures just described can be used also to implement quadratic memories and can be generalized for higher order memories. The basic idea behind this scheme is illustrated in Figure 10 which shows the interconnection between the ith and jth neurons whose weight w_{ij} is a linear combination of all of the inputs and is described by

$$w_{ij} = \sum_{k=1}^{N} \hat{w}_{ijk} x_k. \tag{37}$$

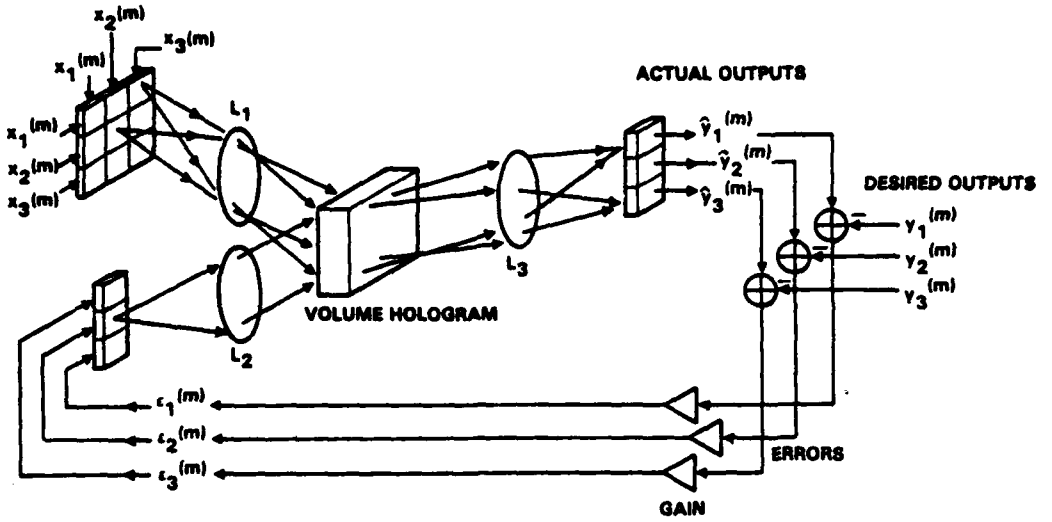


FIGURE 9. Optical system for performing error driven learning in a higher order memory.

The overall result is, of course, recognized to be the equation describing the quadratic memory, but the notion of an input dependent weight suggests the implementation shown in Figure 11. The system is basically an optical vector matrix multiplier (Goodman, Dias, & Woody, 1978) in which the matrix is created on an optically addressed SLM by multiplying the input vector with the three-dimensional tensor stored in a volume hologram. The input vector is represented by a one dimensional array of light sources. The portion of the system on the left side of the SLM is the vector matrix multiplier and it works as follows. Light from each input point is imaged horizontally but spread out vertically so that each source illuminates a narrow, vertical area on the 2-D SLM. The reflectance of the SLM corresponds to the matrix of weights w_{ij} in (37). The reflected light from the SLM travels back towards the input and a portion of it is reflected by a beam splitter and then imaged horizontally but focused vertically onto a 1-D output detector array. The output from the detector array represents the matrix vector product between the input vector and the matrix represented by the 2-D reflectance of the SLM. The matrix of weights, in this case, is not fixed but rather computed from the input via a volume hologram by exposing the righthand side of the SLM as shown in the figure. The optical system to the right of the 2-D SLM in Figure 11 is the same as the $N \mapsto N^2$ system of Figure 8b. The volume hologram which has been prepared to perform the appropriate dimension increasing operation ($N \mapsto N^2$), transforms the light distribution given by its one dimensional array of sources into the input dependent matrix of weights given by (37). This system is functionally equivalent to the previous system except it does not require the use of a 2-D electronically addressed input SLM. The 1-D devices utilized in this architecture are easier and faster to use in practice. Instead a 2-D optically addressed SLM is needed which in practice is simpler to use compared to electronically addressed devices (requires less electronics), typically has more pixels, and is potentially much higher speed. A disadvantage of this method, however, is that it does not lend itself for the direct implementation of the simple outer product training method without the use of an electronically addressed 2-D SLM.

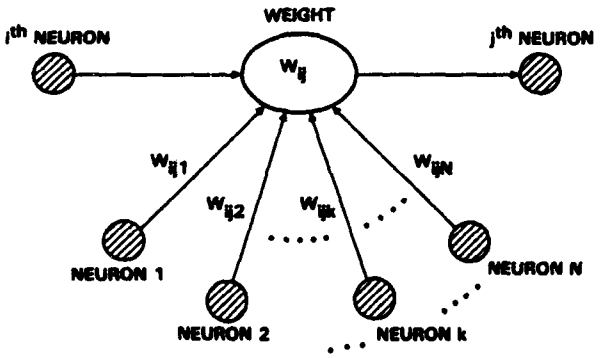


FIGURE 10. Quadratic mappings implemented as nonlinear interconnections.

The $N \mapsto N^2$ mapping technique can be used in conjunction with its inverse, the $N^2 \mapsto N$ mapping, to implement the quadratic outer product memory using two volume holograms, a 1-D electronically addressed SLM, and an optically addressed 2-D SLM. Shown in Figure 12 is a schematic diagram of such a system. The first hologram is prepared with the multiple exposure scheme discussed earlier (Figure 8a) where for each exposure, a memory vector in the one-dimensional input array and one point in the two-dimensional ($\sqrt{M} \times \sqrt{M}$) input training array are turned on simultaneously. The second hologram is prepared by a similar procedure except that the associated output vectors are

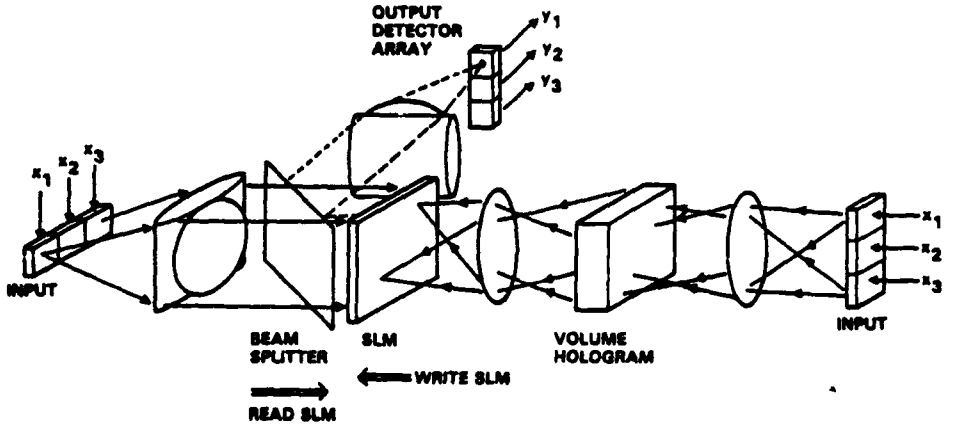


FIGURE 11. Optical architecture for the implementation of the nonlinear interconnections of Figure 10.

recorded in correspondence to each point in the two dimensional training plane. After the holograms are thus prepared, an input vector is loaded into the one-dimensional input array and the correlations between it and the M memory vectors are displayed in the output plane (Athale, Szu, & Friedlander, 1986; Owechko, Dunning, Marom, & Soffer, 1987; Paek & Psaltis, 1987). An optically addressed SLM can be used to produce an amplitude distribution which is the square of the incident correlation amplitudes. The processed light then illuminates the second hologram which serves as an $M \rightarrow N$ interconnection, each correlation peak in the SLM plane reading out its corresponding memory vector and forming a weighted sum of the stored memories on the one dimensional output detector array. This is a direct optical implementation of the system shown in block diagram form in Figure 5 with the 2-D SLM performing the square law nonlinearity at the middle plane and the two-volume holograms providing the interconnections to the input and output.

4.2 Planar Hologram Systems

While not having the extra dimension to directly implement the three dimensional interconnection tensor for general quadratic memories, planar holograms can nevertheless implement the outer product quadratic memory in a way similar to the one used in the system just described. The planar holographic system is shown in Figure 13. Here, the information is stored in the two multichannel 1-D Fourier transform (FT) holograms, the first of which contains the 1-D FTs of the M memory input vectors and the other, the FTs of the associated output vectors (Psaltis & Hong, 1987). The first part of the system is a multichannel correlator which correlates the input against each of the M memory vectors. At the correlation plane, the M correlation functions stacked up vertically are sampled at $x \approx 0$ with a slit to obtain the required inner products which are then squared by the SLM. Each resulting point source of light is then collimated horizontally and imaged vertically onto the second hologram to illuminate that portion which contains the corresponding output vector. The final stage computes the FT of the light distribution just following the second hologram to produce the weighted sum of the vectors at the output detector array. It is interesting to note that if the SLM is removed from the correlation plane, this system reduces to the linear outer product memory.

Notice that in this system if the input pattern shifts horizontally then the correlation peak also shifts in the correlation plane and it is blocked by the slit that is placed there. Therefore shifted versions of the input vector are not recognized, as expected. Shift invariance where the shifted versions of the memory vectors are recognized and their associated outputs, shifted by the

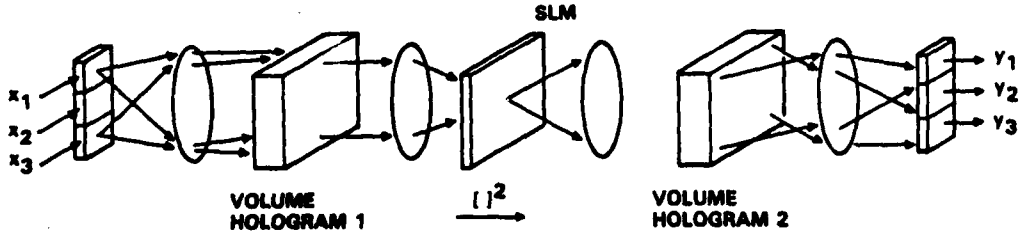


FIGURE 12. Optical higher order associative memory implemented with volume holograms.

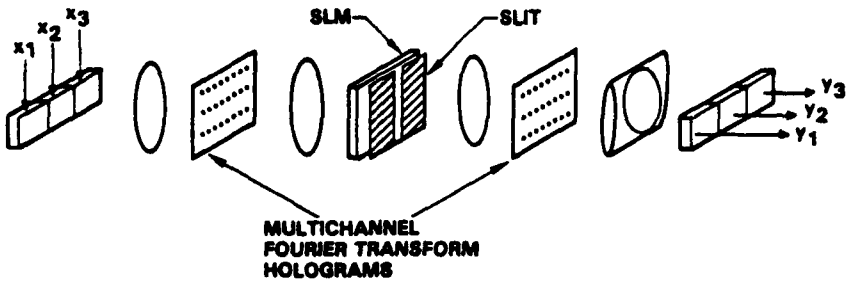


FIGURE 13. Optical implementation of the outer product higher order memory.

same amount as the input, are retrieved can be built into this system by simply lengthening the input SLM and the output detector array to accommodate the shifts and removing the slit in the correlation plane. The resulting system treats each of the $2N - 1$ shifted versions of the memory vectors as a new memory and as a result, the increased capacity of the quadratic memory over the linear one (by a factor of N) is expended to provide invariant operation.

REFERENCES

Abu-Mostafa, Y., & Psaltis, D. (1985). Computation power of parallelism in optical architectures. In *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management* (p. 42). Silver Spring, MD: IEEE Computer Society Press.

Anderson, J. A. (1983). Cognitive and psychological computation with neural models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, 799.

Athale, R. A., Szu, H. H., & Friedlander, C. B. (1986). Optical implementation of associative memory with controlled nonlinearity in the correlation domain. *Optics Letters*, 11(7), 482.

Baldi, P., & Venkatesh, S. S. (1987). Number of stable points for spin-glasses and neural networks of higher orders. *Physics Review Letters*, 58(9), 913.

Chen, H. H., Lee, Y. C., Maxwell, T., Sun, G. Z., Lee, H. Y., & Giles, C. L. (1986). High order correlation model for associative memory. In J. Denker (Ed.), *AIP Conference Proceedings* (p. 86). New York: American Institute of Physics.

Cohen, M., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, 815.

Collier, R. J., Burkhardt, C. B., & Lin, L. H. (1971). *Optical holography*. New York: Academic Press.

Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14, 326.

Duda, R., & Hart, P. (1973). *Pattern classification and scene analysis*. New York: Wiley.

Giles, C. L., & Maxwell, T. (1987). Learning and generalization in higher order networks. *Applied Optics*, 26(23), 4972.

Goodman, J. W., Dias, R. A., & Woody, L. M. (1978). Fully parallel, high speed incoherent optical method for performing discrete Fourier transforms. *Optics Letters*, 2(1), 1.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 2554.

Kogelnik, H. (1969). Coupled theory for thick hologram gratings. *Bell Systems Technical Journal*, 48, 2909.

Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer Verlag.

Maxwell, T., Giles, C. L., Lee, Y. C., & Chen, H. H. (1986). Nonlinear dynamics of artificial neural systems. In J. Denker (Ed.), *AIP Conference Proceedings* (p. 299). New York: American Institute of Physics.

Nakano, K. (1972). Association—A model of associative memory. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2, 380–388.

Newman, C. M. (1987, November). *Memory capacity in symmetric neural networks: Rigorous bounds*. Paper presented at the IEEE Conference on "Neural Information Processing Systems—Natural and Synthetic," Denver, CO.

Owechko, Y., Dunning, G. J., Marom, E., & Soffer, B. H. (1987). Holographic associative memory with nonlinearities in the correlation domain. *Applied Optics*, 26(10), 1900.

Paek, E. G., & Psaltis, D. (1987). Optical associative memory using Fourier transform holograms. *Optical Engineering*, 26(5), 428.

Poggio, T. (1975). On optimal nonlinear associative recall. *Biological Cybernetics*, 19, 201.

Psaltis, D., Brady, D., & Wagner, K. (in press). Adaptive optical networks using photorefractive crystals. *Applied Optics*.

Psaltis, D., & Hong, J. (1987). Shift-invariant optical associative memories. *Optical Engineering*, 26(1), 10.

Psaltis, D., & Park, C. H. (1986). Nonlinear discriminant functions and associative memories. In J. Denker (Ed.), *AIP Conference Proceedings* (p. 370). New York: American Institute of Physics.

Psaltis, D., Park, C. H., & Hong, J. (1986). Quadratic optical associative memories. *Journal of the Optical Society of America—A*, 3(13), 32.

Psaltis, D., Yu, J., Gu, X. G., & Lee, H. (1987). Optical neural nets implemented with volume holograms. In *Proceedings of OSA Second Topical Meeting on Optical Computing* (p. 129). Incline Village, NV: Optical Society of America.

Sejnowski, T. (1986). High-order Boltzmann machines. In J. Denker (Ed.), *AIP Conference Proceedings* (p. 398). New York: American Institute of Physics.

Slepian, D. (1956). A class of binary signaling alphabets. *Bell Systems Technical Journal*, 35, 203.

Van Heerden, P. J. (1963). Theory of optical information storage in solids. *Applied Optics*, 2(4), 393.

Venkatesh, S. S., & Psaltis, D. (in press). Linear and logarithmic capacities of associative memories. *IEEE Transactions on Information Theory*.

Warde, C., & Fisher, A. D. (1987). Spatial light modulators: Applications and functional capabilities. In J. L. Horner (Ed.), *Optical signal processing* (p. 477). San Diego: Academic Press.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *IRE Wescon Convention Record, Pt. 4*, 96.

LIST OF PUBLICATIONS

FIRST PHASE

[1] D. Psaltis, "Two-dimensional optical processing using one-dimensional input devices", IEEE Proc., **72**, 962(1984).

[2] D. Psaltis, "Optical image correlation using acoustooptic and charge-coupled devices", Appl. Opt., **21**, 491(1982).

[3] D. Psaltis, E. Paek, S. Venkatesh, "Acoustooptic/CCD image processor", Proc. Int. Opt. Conf., 204(1983).

[4] D. Psaltis, "Incoherent electoroptic image correlator", Opt. Eng., **23**, 698(1984).

[5] D. Psaltis, E. Paek, and S. Venkatesh, "Optical Image Correlation with a Binary Spatial Light Modulator", Opt. Eng. , **23**, 698(1984).

[6] D. Psaltis, F. Mok and E. G. Paek, "On the Use of the Litton Magneto-Optic Device in Optical Processors", SPIE Proc., **465**, (1984).

[7] Kelvin Wagner and D. Psaltis, "Real Time Computation of Moments with Acousto-optics", SPIE Proc., **352-19**, (1982).

[8] Yaser Abu-Mostafa and D. Psaltis , "Recognitive Aspects of Moment Invariants", IEEE Trans. Pattern Analysis and Machine Intelligence, **PAMI-6**, 698(1984).

[9] Yaser Abu-Mostafa and D. Psaltis, "Image Normalization by Complex Moments", IEEE Conf. Computer Vision and Pattern Recognition, p. 114, June 1983.

[10] Y. Shi, D. Psaltis, A. Marrakchi, and A. Tanguay, "Photorefractive Incoherent to Coherent Optical Converter", Appl. Opt., **22**, 3665(1983).

[11] A. Marrakchi, A.R. Tanguary, Jr., J. Yu, and D. Psaltis, "Physical Characterization of the Photorefractive Incoherent to Coherent Optical Converter", Opt. Eng., **24**, 124(1985).

[12] D. Psaltis and Jeff Yu, "Photorefractive Incoherent to Coherent Optical Conversion", SPIE Proc., **465-01**, (1984).

[13] A. Marrakchi, A. Tanguay, J. Yu, and D. Psaltis, "Photorefractive Incoherent to Coherent Converter: Materials Issues" SPIE Proc., **465-12**, (1984).

[14] M. Haney and D. Psaltis, "Measurement of the Temporal Coherence Properties of Pulsed Single-Mode Laser Diodes" Appl. Opt., **24**, 1926(1985).

[15] N. Farhat, D. Psaltis, A. Prata, and E. G. Paek, "Optical Implementation of the Hopfield Model", Appl. Opt., **24**, 1469(1985).

SECOND PHASE

[16] D. Psaltis, J. Yu, and J. Hong, "Bias-free time-integrating optical correlator using a

photorefractive crystal", Appl. Opt., 24, 3860(1985).
[17] E. G. Paek, C. H. Park, F. Mok, and D. Psaltis, "Acoustooptic image correlators", SPIE Proc. 638-05, (1986).
[18] E. G. Paek and D. Psaltis, "Optical Associative Memory using Fourier Transform Holograms", Opt. Eng., 26, 428(1987).
[19] D. Psaltis, C. H. Park, and J. Hong, "Higher order associative memories and their optical implementations", Neural Networks, 1, 149(1988).
[20] Ken Hsu, David Brady, and Demetri Psaltis, *Experimental demonstrations of optical neural computers*, Proc. IEEE conf. on Neural Information Processing Systems, Denver, November 1987.
[21] S. Hudson and D. Psaltis, "Optical network that learns to perform motion compensation in radar imaging", SPIE Proc., 882, 154(1988).
[22] K. Wagner and D. Psaltis, "Multilayer optical learning networks", Appl. Opt., 26, 5061(1987).
[23] D. Psaltis, D. Brady, X. G. Gu, and K. Hsu, "Optical implementation of neural computers", **Optical Processing and Computing**, H. Arsenault, ed., Academic Press, (New York)1989.
[24] K. Hsu and D. Psaltis, "Invariance and discrimination properties of the optical associative loop", Proc. IEEE conf on Neural Networks, San Diego, July 1988.
[25] D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals", Appl. Opt., 27, 1752(1988).
[26] H. Lee, X. G. Gu, and D. Psaltis, "Volume holographic interconnections with maximal capacity and minimal cross talk", To appear in J. Appl. Phys., (1989).
[27] D. Psaltis, M. Neifeld, and A. Yamamura, "Optical disk based correlation architectures", to appear in Opt. Let., (1989).
[28] D. Psaltis, A. Yamamura, and M. Neifeld, "Parallel readout of optical disks", Optical Computing, 1989 Technical Digest Series, 9 (Optical Society of America, Washington, D.C. 1989).

END

FILMED

7-89

DTIC