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.
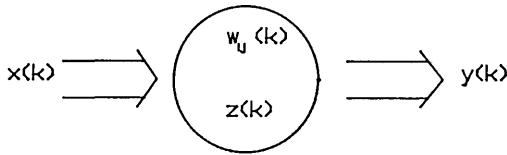
Fig. 1. General neural network architecture.

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

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

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:

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

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:

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

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 sys-

tems, however, this type of learning can provide an effective mechanism for matching the available resources (connections and neurons) to the requirements of the problem. In optical realizations error driven algorithms are more difficult to implement than prescribed approaches due to the need for dynamically modifiable interconnections and the incorporation of an optical system that monitors the performance and causes the necessary changes in the weights.[26] While this problem could be avoided by performing learning off line in computer simulations and recording the optimized interconnection matrix as in prescribed learning, this approach has the disadvantage that once again the matrix is fixed a priori, thus preventing the network from being adaptive. In subsequent sections we will consider a relatively simple form of Eq. (3) in which $\Delta w_{ij}(k)$ depends only on locally available information, i.e., z_i in one layer and z_j in an adjacent layer:

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

The perceptron and the backward error propagation algorithms both fall in this subcategory if we allow the neuronal activity z_i to include error signals, i.e., if each neuron has distinct signal and error outputs which are separated temporally or spatially. An example of such a neuron implemented in optics is given below in conjunction with an optical back error propagation system.

In the case of self-organizing learning algorithms we require not that the specified inputs produce a particular response but rather that they satisfy a general restriction, often imposed by the structure of the network itself. Since there is no a priori expected response, the learning rule for self-organizing systems is simply

$$\Delta w_{ij}(k) = f_{ij}[\mathbf{x}^{(m)}, w_{rs}(k)]. \qquad (5)$$

This type of learning procedure can be useful, for example, at intermediate levels of a network where the purpose is not to elicit an external response but rather to generate appropriate internal representations of the information that is presented as input to the network. There is a broad range of self-organizing algorithms, the simplest of which is probably lateral inhibition to enforce grandmother cell representations.[10,27] The objective of the learning procedure is to have each distinct pattern in an input set of neurons activate a single neuron in a second set. In the architecture shown in Fig. 2 this is accomplished via inhibitory connections between the neurons in the second set. Once a particular neuron in the second layer is partially turned on for a specific pattern it prevents the connections to the other neurons in the second set from assuming values that will result in activity at more than one neuron. The details of the dynamics of such procedures can be quite complex (e.g., see Ref. 28), as can corresponding optical implementations. An advantageous feature of optics in connection with self-organization is that global training signals, such as fixed lateral inhibition between all the neurons in a given layer, can easily be broadcast with optical beams.
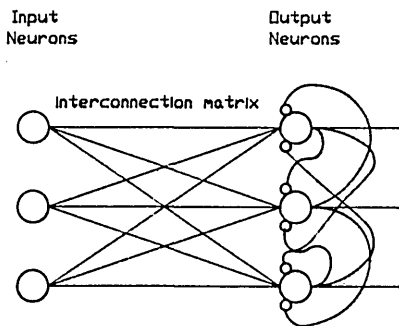
Fig. 2. Two-layer network with lateral inhibition. Connections ending with an open circle are inhibitory.

Fig. 3. Optical neural computer architecture.

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

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

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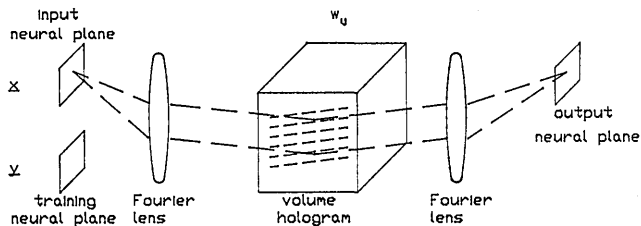
in the volume medium. This grating diffracts an input beam at spatial frequency $\mathbf{k}_\alpha$ into an output beam at spatial frequency $\mathbf{k}_\beta$ if these two beams satisfy the Bragg constraint that

$$\mathbf{k}_\alpha - \mathbf{k}_\beta = \mathbf{K}_{ij}. \qquad (7)$$

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

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

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

$$\nu_i = 0, \pm 1, \pm 2 \ldots, \qquad (9)$$

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

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

between the field amplitude, $E_{j\,\text{out}} \exp(j\mathbf{k}_j \cdot \mathbf{r})$, of the jth neuron and the field amplitude, $E_{i\,\text{in}} \exp(j\mathbf{k}_i \cdot \mathbf{r})$, 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

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

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

$$\eta_1 \approx \frac{\eta_0}{\sqrt{\dfrac{N}{2}}}. \quad (12)$$

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 $E_{j\,\text{out}} \exp(j\phi_j)$ 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

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

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 $|E_{j\,\text{out}}|^2$ and w_{ij} is represented by η_{ij}^2. The transformation between $\{x_j\}$ and $\{y_i\}$ takes the form

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

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

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

From Eq. (12) we find

$$G_\text{coherent} = \frac{1}{2\eta_0}. \quad (16)$$

For an incoherent system the corresponding relationship is

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

In this case Eq. (12) yields

$$G_{\text{incoherent}} = \frac{\sqrt{N}}{2\eta_0}. \qquad (18)$$

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, $\eta_0 = 0(1)$. This means that in coherent systems relatively little gain [i.e., $G = 0(1)$] 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,

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

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

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

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),

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

or

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

By induction, for $m > m_0$

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

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.
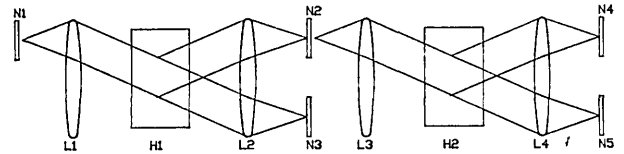

Fig. 4. Optical architecture for backward error propagation learning.

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

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

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

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

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
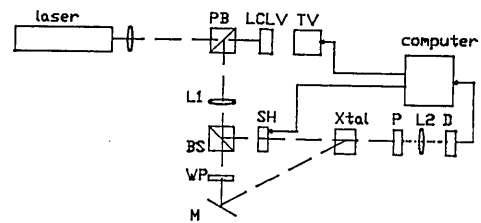

Fig. 5. Simple photorefractive learning system: PB is a polarizing beam splitter; $L1$ and $L2$ are imaging lenses; WP is a quarterwave plate; SH is a shutter; P is a polarizer; D is a detector; M is a mirror.
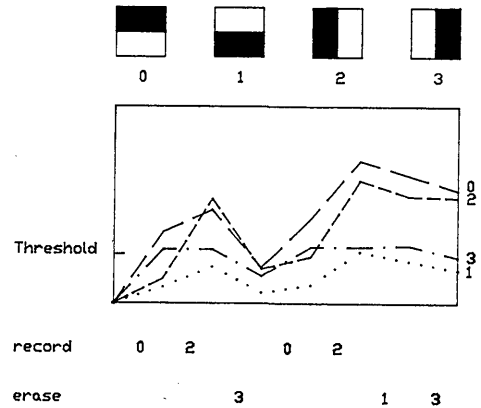

Fig. 6. Experimental learning curves.

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.

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).