

# Programmable image associative memory using an optical disk and a photorefractive crystal

Mark A. Neifeld and Demetri Psaltis

The optical disk is a computer-addressable binary storage medium with very high capacity. More than  $10^{10}$  bits of information can be recorded on a 12-cm-diameter optical disk. The natural two-dimensional format of the data recorded on an optical disk makes this medium particularly attractive for the storage of images and holograms, while parallel access provides a convenient mechanism through which such data may be retrieved. In this paper we discuss a closed-loop optical associative memory based on the optical disk. This system incorporates image correlation, using photorefractive media to compute the best association in a shift-invariant fashion. When presented with a partial or noisy version of one of the images stored on the optical disk, the optical system evolves to a stable state in which those stored images that best match the input are temporally locked in the loop.

*Key words:* Associative memory, optical disk, image correlation, photorefractive media.

## 1. Introduction

The retrieval of information from planar optical storage media (e.g., optical disks) by using parallel access offers several advantages over the conventional serial-access mode of retrieval. Perhaps the most notable among these is the large potential increase in data transfer rate between mass storage and processors achievable with parallel access. Implementing only one-dimensional (1-D) across-track parallelism with 1000 output detectors yields an increase in data transfer rate from a conventional 40 Mbits/s to 40 Gbits/s (before error-correction decoding).<sup>1</sup> This is a particularly attractive prospect in light of the fact that the conventional serial channel between mass storage and processor represents a severe bottleneck in many applications. In addition to reducing the memory interface bottleneck present in conventional computing architectures, parallel access offers the capability for realizing a number of specialized processors as well. Many of these specialized processors have been discussed elsewhere and include data base machines, neural networks, and image correlators.<sup>2,3</sup> In this paper we describe the use of an optical-disk-based image correlator in the

realization of a correlation-based image associative memory.

The correlation-based associative memory (AM) dates back to Gabor and is based on the notion that the correlation signal can be used as a metric to compare an unknown input image with a library of programmed reference images in a shift-invariant fashion.<sup>4</sup> Having determined the closest match, one can retrieve this reference image from the system. Such a system therefore operates as an autoassociative memory. A number of researchers have extended this basic idea to arrive at a variety of interesting optical architectures.<sup>5-7</sup> In comparison with these previous systems, the principal attractions of the present system are its full two-dimensional (2-D) shift invariance and its relatively near-term practicality. Since the system contains only mature optical and electronic technologies, its realization is straightforward. Additional characteristics of the present system include real-time programmability, guaranteed convergence in constant time, good light efficiency, and large image capacity. These features together with the shift-invariance characteristic are further described in later sections.

This paper is divided into four sections. Section 1 describes in detail the photorefractive volume holographic image correlator that utilizes an optical disk. With disk rotation, 2-D shift invariance is recovered in this system. This section also discusses the light efficiency of the resulting implementation. Section 2 describes the closed-loop optical-disk-based associative memory. A discussion of the electronic process-

---

M. A. Neifeld is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721. D. Psaltis is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125.

Received 25 February 1992.

0003-6935/93/234398-12\$06.00/0.

© 1993 Optical Society of America.

ing required to close the memory loop is included along with experimental results. Section 3 addresses the effect of disk imperfections on image correlators and is followed by conclusions in Section 4.

## 2. Optical-Disk-Based Photorefractive Correlator

### A. System Description

A schematic of the photorefractive (PR) correlator is shown in Fig. 1. It has been described previously in Ref. 2. We briefly outline the operation of the correlator here for clarity. In the system shown in Fig. 1 a photorefractive crystal is used to record a hologram of the input Fourier transform. In place of the PR crystal, any other real-time holographic storage medium may be used; however, the PR approach was chosen here to take advantage of the high diffraction efficiency obtainable from volume holograms while retaining the real-time nature of the recording process. During the recording phase the disk illumination is blocked and the input transparency is illuminated from the right. For real-time applications this input transparency may be a spatial light modulator (SLM) interfaced to a video camera, for example. A hologram is formed in the crystal between the input Fourier transform and the reference beam, as shown. The hologram thus recorded in the PR crystal is read out by using the reference library from the disk. On readout, the input is blocked and the disk is illuminated. The product of the input and reference Fourier transforms is formed in the crystal, and the diffracted light is inverse transformed to obtain the desired correlation pattern in the output plane. Owing to the volume nature of the hologram stored in the PR crystal, only a single vertical slice of the 2-D correlation pattern is actually obtained in the output plane of Fig. 1. This plays an

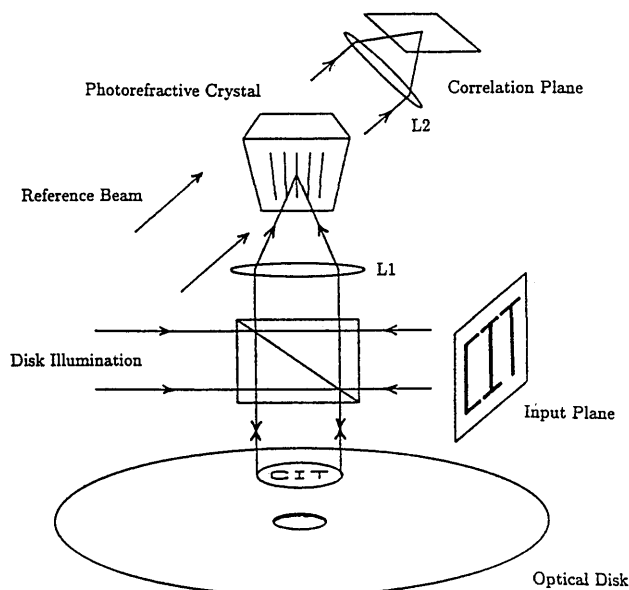


Fig. 1. Schematic of the photorefractive optical-disk-based image correlator.

important role in the behavior of the PR correlator, as we discuss in Subsection 2.B.

In order for a conventional optical disk to support image storage in the fashion depicted in Fig. 1, we require that the disk pixels exhibit across-track coherence. This characteristic of the disk storage format provides a coherent polar grid of pixels on which we may record image data. Although not a common capability of commercial disk systems, across-track coherence is a characteristic of the Sony sampled format disk system that we used in our experiments. Other considerations regarding the storage of image data on the optical disk, including the issue of area usage and the effects of track curvature, have been discussed in detail in Ref. 1 and are not addressed here.

### B. Shift-Invariance Recovery

If the photorefractive crystal in Fig. 1 is replaced by a thin medium such as a holographic plate (i.e., to realize a simple VanderLugt correlator), then the output pattern is exactly the desired 2-D correlation.<sup>8</sup> When a thick hologram is used in the filter plane of such a system, however, the resulting output is a 1-D slice of the 2-D correlation pattern.<sup>9</sup> This can be simply understood as follows. On recording, each plane wave corresponding to each of the points in the input image forms a grating with the reference beam. The resulting gratings exhibit Bragg selectivity in the along-track direction. As a result, on readout a point along a given radial line on the disk can only read out those gratings formed by points along one vertical line in the input. Each such line on the disk reads out a corresponding set of gratings and generates a vertical array of spots in the correlation plane at the horizontal location corresponding to the reference beam Fourier-transform position. The coherent sum of all such reconstructions comprises the output of the correlator. This output pattern is the desired 2-D correlation multiplied, in the horizontal direction, by a sinc function whose width is inversely proportional to the hologram thickness. In the system of Fig. 1 this property does not represent a limitation since all 1-D slices are obtained sequentially by means of disk rotation. Furthermore, instead of a full 2-D detector array being required at the output, a 1-D array is sufficient to sequentially detect each slice of correlation output. This represents a potential advantage over the VanderLugt system since only  $N$  detectors are required to operate at 40 MHz as opposed to  $N^2$ , thereby reducing the readout channel bandwidth to 4 GHz for  $100 \times 100$  pixel reference images and a 40-Hz disk rotation rate. It should be pointed out that for some applications the time cost associated with sequential correlation-plane readout represents a serious limitation; however, for video-rate applications this should not be the case. The hologram thickness ( $L$ ) can be increased to improve diffraction efficiency. However, since the crystal thickness also results in Bragg selectivity in the across-track direction, making  $L$  too large limits vertical shift invariance as well. We can derive a

condition such that both of these constraints may be satisfied. Specifically, we want to derive the PR crystal thickness required to achieve one line of correlation in the horizontal dimension while retaining  $N$  pixels of shift invariance in the vertical.

The condition for no shift invariance in the horizontal or  $x$  direction is understood by considering Fig. 2, in which we show two input points located at  $y = 0$  and separated by  $\Delta x$ . This figure represents two adjacent points in the input plane of the PR correlator. In the Fourier plane each input point gives rise to a plane wave, which may be characterized by the  $x$  component of its propagation vector. Specifically,

$$k_x^1 = \pi\Delta x / \lambda F, \quad (1)$$

$$k_x^2 = -\pi\Delta x / \lambda F, \quad (2)$$

where  $F$  is the Fourier-transform lens focal length,  $\lambda$  is the system operating wavelength, and the superscripts indicate the input point with which the plane wave is associated. In order for the plane wave arising from point 1 to read out the grating formed by the interference between point 2 and the reference, from the Bragg condition we require that  $\Delta k_x < 2/L$ , where  $L$  is the grating thickness and  $\Delta k_x = (k_x^1 - k_x^2)$ . Conversely, in order to guarantee that points in adjacent columns do not read out each other's gratings, we require that  $\Delta k_x > 2/L$ . Substituting  $k_x^1$  and  $k_x^2$  from above, we see that in order for exactly one line of correlation to appear in the output plane of the PR correlator, the crystal thickness must satisfy  $L > \lambda F / \pi\Delta x$ . This inequality states that smaller input-plane resolution requires larger crystal thickness to eliminate cross talk. For reasonable system parameters such as  $F = 10$  cm,  $\lambda = 633$  nm, and  $\Delta x = 10$   $\mu$ m, we have  $L > 2$  mm. This is a reasonable requirement since, in general, large  $L$  is desirable owing to the corresponding increase in diffraction efficiency resulting from the long interaction length. There is a maximum desirable crystal thickness, however, since for very long interaction length, shift invariance may be lost in the  $y$  direction as well. In order to better understand the following discussion, consider Fig. 3. The recording geometry is depicted in Fig. 3(a) and resembles the geometry used above,

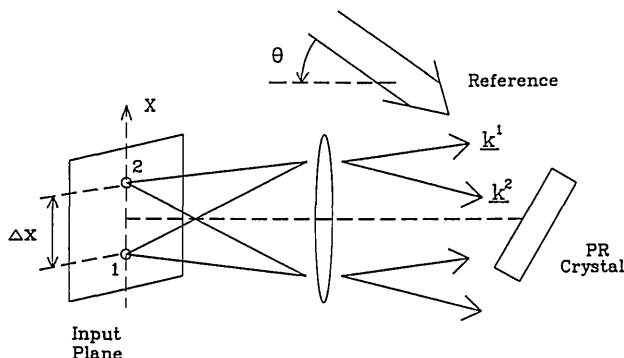


Fig. 2. Geometry used for calculating horizontal shift invariance.

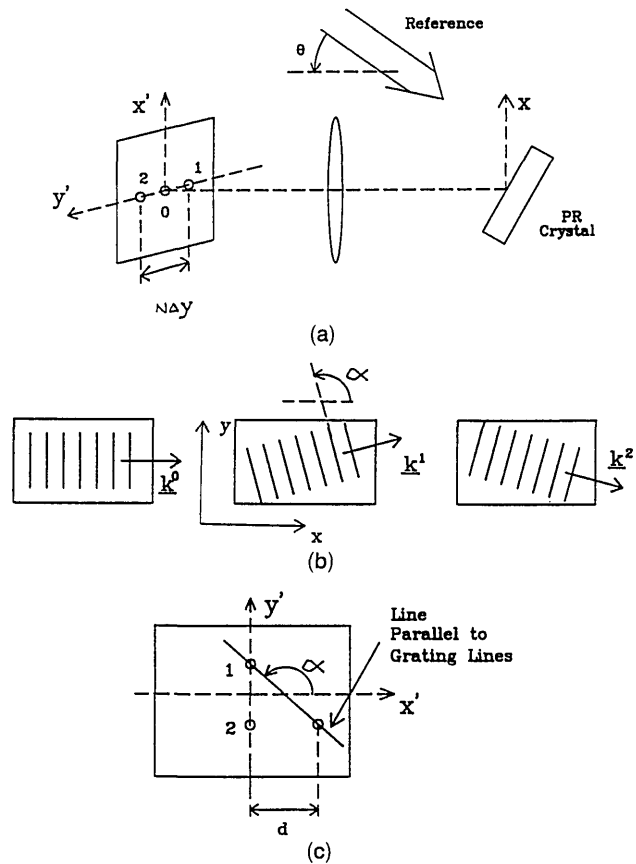


Fig. 3. Geometry used for calculating vertical shift invariance.

but the input plane now consists of three points along the  $y$  axis at  $y = 0$  and  $y = \pm N\Delta y/2$ . As before, each point gives rise to a plane wave that interferes with the reference beam. The four interfering fields of interest are given as

$$E_{\text{REF}} \propto \exp[i(kx \sin \theta + kz \cos \theta)] \approx \exp(ikx \sin \theta) \exp(ikz), \quad (3)$$

$$E_0 \propto \exp(ikz), \quad (4)$$

$$E_1 \propto \exp(ikyN\Delta y/2F) \exp(ikz), \quad (5)$$

$$E_2 \propto \exp(-ikyN\Delta y/2F) \exp(ikz), \quad (6)$$

where  $\Delta y$  is the input-plane resolution in the  $y$  direction,  $N$  is the number of pixels in a single column of the input, and  $k = 2\pi/\lambda$ ; we make the simplifying assumption that the medium in which the interference takes place is isotropic with unity refractive index. Recently a more complete analysis has been performed and yields the same results.<sup>10</sup> The three gratings arising from these four fields are shown in Fig. 3(b). As an example, we consider the second grating of Fig. 3(b) and write an expression for one of the grating lines or planes of constant phase:

$$\frac{N\Delta y}{2F} y + x \sin \theta = 0. \quad (7)$$

From this equation we can determine the angle  $\alpha$  shown in the figure as  $\alpha = \tan^{-1}(2F \sin \theta / N\Delta y)$ . If we now go back to the input plane shown in Fig. 3(c), it is clear that moving away from point 1 along the line shown does not result in Bragg mismatch since we are moving along a line parallel to the grating lines. Specially, we consider a point at the bottom of the image plane lying along this line; it will be able to read out the grating from point 1. We want point 2 to be able to read this grating. This requires that  $d < \Delta x_{\text{MAX}}$ , where  $d$  is the distance shown and  $\Delta x_{\text{MAX}}$  is the maximum shift in the  $x$  direction that can be tolerated as determined by the crystal thickness  $L$ .  $\Delta x_{\text{MAX}}$  can be determined from our previous result as  $\Delta x_{\text{MAX}} = \lambda F / \pi L$ . Putting these conditions together with our expression for  $\alpha$ , we have  $d < \Delta x_{\text{MAX}}$  and  $d = N\Delta y / \tan(\alpha)$ , which determines an upper bound on the crystal thickness as

$$L < 2\lambda F^2 \sin \theta / \pi(N\Delta y)^2. \quad (8)$$

For typical parameters such as  $\Delta x = \Delta y = 10 \mu\text{m}$ ,  $N = 100$ , and  $\theta \approx 60^\circ$  we obtain  $L < 3.5 \text{ mm}$ . Once again we see that this thickness is in a reasonable range of available crystal thicknesses and is also compatible with the previous result. Summarizing these two results, we see that in order for the PR correlator to generate a single line of correlation in the output plane while not sacrificing shift invariance in the orthogonal dimension, we require that

$$\lambda F / \pi \Delta x < L < 2\lambda F^2 \sin \theta / \pi Y^2, \quad (9)$$

where  $Y$  is the vertical extent of the input image. It is interesting to see when such a crystal thickness exists. We find that  $1/\sin \theta < 2F/\text{SBP}_{\text{IN}}\Delta x$ , where we set  $\Delta x = \Delta y$  and label  $N^2 = \text{SBP}_{\text{IN}}$  as the space-bandwidth product of the input plane equal to the total number of input pixels.

### C. Photorefractive Correlator Efficiency

We would like to compute the amount of light reaching the autocorrelation peak in the PR correlator. In the case of Fourier-based correlation the autocorrelation peak power  $P_c$  can be calculated by using a simple statistical approach. In order to compute  $P_c$  we first compute the expected fraction of the diffracted power that arrives at the peak of the autocorrelation plane. We assume that we have random images and consider two cases. First we treat the case of bipolar images. Let  $a_{ij} \in \{\pm 1\}$  represent the value of the  $(i, j)$  pixel in the random image, and let  $a_{ij}$  be characterized by the simple statistics

$$p(a_{ij} = 1) = p(a_{ij} = -1) = 1/2, \quad \forall i, j, \quad (10)$$

$$p(a_{ij} = k, a_{mn} = k') = p(a_{ij} = k)p(a_{mn} = k'), \quad \forall i \neq m, j \neq n, \quad (11)$$

where  $p(a_{ij} = k)$  is the probability that the  $(i, j)$  pixel takes on the value  $k$ . The second condition denotes statistical independence among pixels. For compari-

son purposes we first use these image characteristics to write an expression for the value of the electric field at each point in the correlation plane of the VanderLugt system ( $c_{mn}$ ) as

$$c_{mn} = \sum_{i=m+1}^N \sum_{j=n+1}^N a_{ij} a_{i-m, j-n}, \quad (12)$$

$$c_{00} = \sum_{i=1}^N \sum_{j=1}^N (a_{ij})^2, \quad (13)$$

where we ignore unimportant proportionality constants and write  $c_{mn}$  as the value of the 2-D correlation pattern at the  $(m, n)$  output pixel. For bipolar images,  $c_{00} = N^2$ ; therefore the optical power diffracted into the autocorrelation peak is  $c_{00}^2 = N^4$ . In order to find the expected power away from the peak, we must first calculate the expected power measured at  $(m, n)$ , which is

$$\begin{aligned} E\{c_{mn}^2\} &= E\left\{ \sum_{i=m+1}^N \sum_{i'=m+1}^N \sum_{j=n+1}^N \sum_{j'=n+1}^N a_{i,j} a_{i-m, j-n} a_{i',j'} a_{i'-m, j'-n} \right\} \\ &= (N-m)(N-n). \end{aligned} \quad (14)$$

The total power diffracted outside the peak therefore is given by

$$P_\Sigma = 4 \sum_{m=1}^N \sum_{n=1}^N E\{c_{mn}^2\}, \quad (15)$$

where the factor of 4 is used to take into account all four quadrants of the correlation plane. Finally we find that the fraction of diffracted power reaching the autocorrelation peak is given by

$$\eta_P = c_{00}^2 / (P_\Sigma + c_{00}^2) \quad (16)$$

$$= N^4 / (2N^4 - 2N^3 + N^2), \quad (17)$$

which for large  $N$  reduces to  $\lim_{N \rightarrow \infty} \eta_P = 1/2$ . This means that one half of the diffracted power from the VanderLugt system reaches the peak when the input and reference images are matched bipolar images. We can perform a similar analysis for the case of unipolar input and references. Without including the details of this computation, we find that in the limit of large  $N$ , we have  $\lim_{N \rightarrow \infty} \eta_P = 1/N^2$  for unipolar images.

In contrast to the VanderLugt system, the PR correlator utilizes a high efficiency ( $\eta > 50\%$ ) volume hologram and computes adjacent lines in the correlation pattern sequentially, thereby improving light efficiency drastically. In order to calculate the efficiency, we return to the bipolar image statistics given above. Once again we can express the autocorrelation peak power as  $P_0 = c_{00}^2$ , where  $c_{00}$  is given in Eq. (13). In the present case, however, the total diffracted power that appears away from the peak is modified as compared with the VanderLugt system.

Since only one line of correlation may be detected at any given time, we can write an expression for the field away from the peak as

$$c_k = \sum_{i=1}^N \sum_{j=k+1}^N \alpha_{ij} \alpha_{i,j-k}, \quad (18)$$

where a single index  $k$  indicates the position along the single line of correlation output. Once again we find the expected value of the power measured at any pixel off peak as  $P_k = E\{c_k^2\}$ , and the total diffracted power off peak has an expected value of

$$P_\Sigma = 2 \sum_{k=1}^N E\{c_k^2\}. \quad (19)$$

Using bipolar image statistics, we find that  $E\{c_k^2\} = N(N - k)$ , yielding  $P_\Sigma = N^3 - N^2$  and  $P_0 = N^4$  so that the fraction of diffracted power reaching the peak is given by Eq. (16), and for large  $N$  we obtain  $\lim_{N \rightarrow \infty} \eta = 1$ . This result reveals that for large enough random bipolar input images, we can expect essentially all of the diffracted light to reach the correlation peak when there is a match between input and reference. In order to verify this result, computer simulations of the PR correlator were performed. The simulation procedure consisted of first choosing a random bipolar image of a particular size. The 2-D autocorrelation of this image was computed and the resulting pattern was squared to represent correlation-plane intensity. The ratio of the peak of this intensity pattern to the sum of all pixel values in the central column is the desired efficiency  $\eta$ . Using image sizes in the range  $2 \times 2$  to  $50 \times 50$  and 100 statistics for each image size, we generated the plots in Fig. 4. As we see from Fig. 4(a) for  $N = 50$ , 98% of the light energy diffracted by the PR hologram reaches the peak. A similar analysis can be performed for unipolar images, and we obtain  $\lim_{N \rightarrow \infty} \eta = 6/N$ . The efficiency of the unipolar PR correlator is higher by a factor of  $N$  than that of the unipolar VanderLugt system. Again, this result is verified by computer simulation, and the result is shown in Fig. 4(b).

Given the efficiencies derived above, we can estimate the peak detectability in the PR correlator. With reference images of  $100 \times 100$  pixels, 10-mW readout illumination, a disk efficiency of  $\eta_D = 0.1\%$ , and a disk-rotation rate of 40 Hz, we obtain a peak dwell of  $t_D = 25$  ns for a 400,000-image/s correlation rate. This corresponds to an autocorrelation peak power of  $5 \mu\text{W}$  for the bipolar system and  $0.3 \mu\text{W}$  for the unipolar case. In the previous calculation we assume a 50% photorefractive diffraction efficiency. We may compute the number of photons in the autocorrelation peak for these two cases by using an illuminating wavelength  $\lambda = 633$  nm. The result is that, for the PR correlator operating on bipolar images, we have a peak detectability of  $n_p = 4 \times 10^5$  photons, and in the unipolar case we obtain  $n_p = 2.4 \times 10^4$  photons, which are easily measurable quantities.

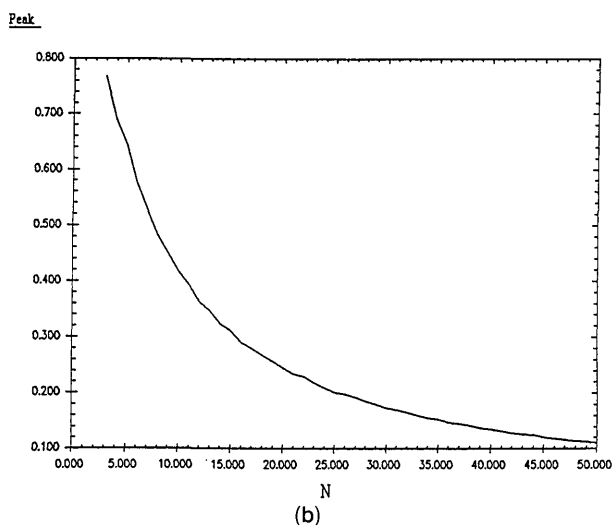
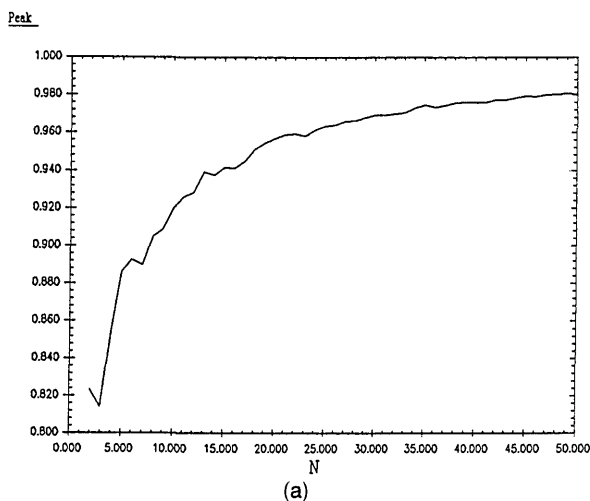


Fig. 4. Photorefractive correlator diffraction efficiency simulation for (a) bipolar and (b) unipolar images.

#### D. Experimental Demonstration of the Photorefractive Correlator

The PR correlator shown in Fig. 1 was experimentally demonstrated. The input to the correlator is shown in Fig. 5 and was recorded in sector 1 of a Sony alloy-mode write-once optical disk; the input measured  $2 \text{ mm} \times 4 \text{ mm}$ . The Sony optical-disk system used in this work has been characterized in Ref. 1. The photo in Fig. 5 was taken from the disk by using the diffractive retrieval scheme discussed in Ref. 1. As an input SLM we also used an optical disk on



Fig. 5. Image used for PR correlator experiments. The photograph was taken from the disk by using the diffractive readout scheme described in Ref. 1.

which we recorded the same image data. The use of the optical disk as an input SLM along with its relatively low diffraction efficiency ( $\eta = 0.1\%$ ) resulted in a saturation exposure time of  $\sim 16$  s for the SBN crystal used in the experiment. We assume negligible birefringence in the disk coating layer and orient both the reading and writing polarizations parallel with the crystal  $c$  axis to take advantage of the large  $r_{33}$  coefficient in SBN ( $r_{33} = 1340 \times 10^{-12}$  m/V). The correlation results obtained with this setup are shown in Figs. 6–8. The acronym CIT was recorded on two different sectors of the same disk for use as input and reference images. There are two distinct configurations for the PR correlator; the configuration in which the Bragg selectivity is in the horizontal or along-track direction and the complementary configuration in which the system is shift invariant in the horizontal direction. The difference between these two arrangements lies in how the reference beam is presented to the crystal. In the former configuration the reference beam is presented so as to produce a principle grating vector that is parallel with the along-track direction on the disk; the later configuration produces a grating vector that is parallel with the radial direction on the disk. In actual practice we would implement the correlator in the former configuration as shown in Fig. 1, which permits disk rotation to recover horizontal shift invariance; however, for the purposes of demonstration both configurations were tested. Results from these two configurations are shown in Figs. 6 and 7. The autocorrelation of the acronym CIT has a five-lobed structure in the  $x$  direction and essentially no structure in the  $y$  direction. In Fig. 6, which has shift invariance in  $x$ , we do indeed see this structure, whereas in Fig. 7, which has Bragg selectivity in  $x$ , we see none. These results were obtained by using identical recording and readout images (i.e., from the same sector). In Fig. 6 we notice that the outermost

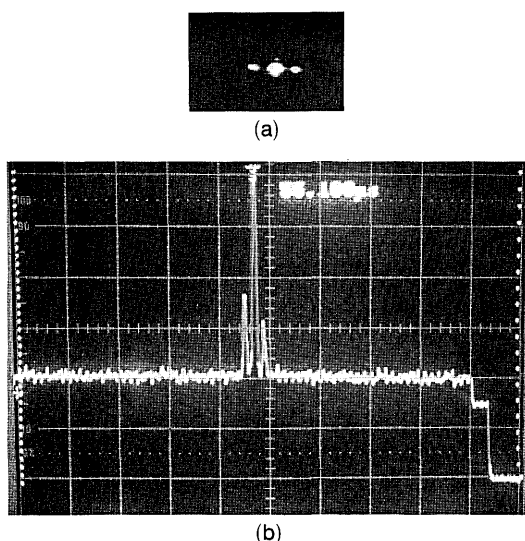


Fig. 6. Correlation results for shift invariance in the horizontal direction: (a) 2-D correlation pattern, (b) slice through  $y = 0$  showing a multilobed pattern.

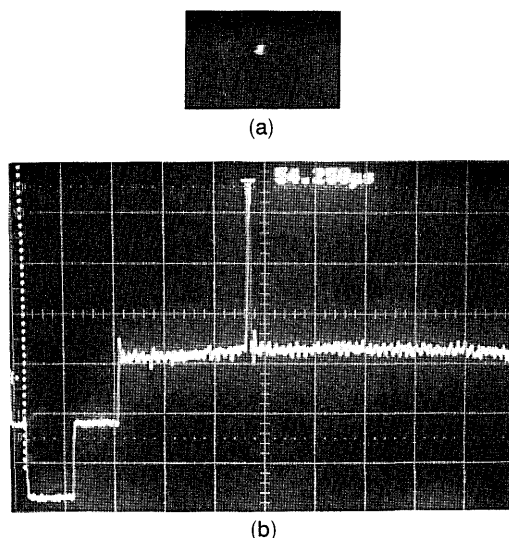


Fig. 7. Correlation results for shift invariance in the vertical direction: (a) 2-D correlation pattern, (b) slice through  $y = 0$  showing a single-lobed pattern.

two sidelobes are missing from the autocorrelation pattern. This is in agreement with our previous result, which predicted a loss in shift invariance in the direction perpendicular to the Bragg direction owing to excessive crystal thickness. Using values from our experimental system, we have  $F = 12$  cm,  $\lambda = 488$  nm,  $L = 5$  mm, and  $\theta = 60^\circ$ , which yields a permissible image extent of  $Y < (2\lambda F^2 \sin \theta / L)^{1/2} = 1.5$  mm. Our actual image size was  $\sim 4$  mm so that we would expect to lose the sidelobes of the correlation pattern. We also obtained correlation signals

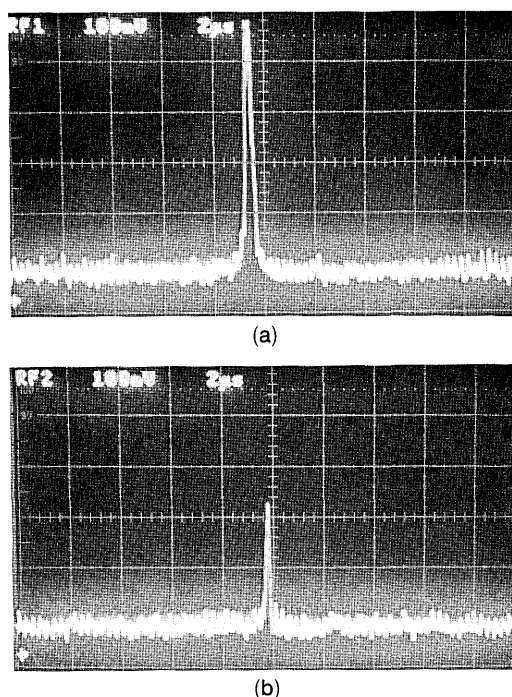


Fig. 8. Correlation results for the input image from sector 1 and reference image from sector 2: (a) expanded correlation peak of Fig. 7(b) for comparison with (b) the reduced signal from sector 2.

from readout, using a different sector image. These results are shown in Fig. 8. For this configuration we observed a significant reduction (3 dB) in signal level. Although phase nonuniformities from disk coating materials and hologram erasure from readout play minor roles in the signal reduction observed, the primary factor is believed to be disk alignment. As the disk is rotated to illuminate the alternate reference image recorded in sector 2, disk wobble and tilt results in a slight shift of the Fourier transform and a corresponding reduction in correlation peak intensity. A more detailed discussion of the sensitivity of the PR correlator to disk imperfections is included in Section 4.

### 3. Closed-Loop Optical-Disk-Based Associative Memory

#### A. System Description

The optical-disk-based associative loop is shown schematically in Fig. 9. The system comprises the PR correlator described earlier with a 1-D detector array followed by some simple electronics that generate the loop feedback, as shown. The system operation is as follows. First, with reference to Fig. 9, the input and reference illumination is on and the disk illumination is off. A Fourier-transform hologram of the input image is formed in the crystal with the reference beam, as before. The readout phase is initiated first by turning the reference and input illumination off and then by illuminating the transmissive disk from below for one rotation. The correlation between the input image and each illuminated image on the optical disk appears at the output plane at some point in time as the disk rotates. In order to measure the sequence of correlation columns that appear in the output plane as the disk rotates, a 1-D detector array is used. By choosing the largest element of the detected signal at any time and then by further choosing the maximum such largest signal over one full disk rotation, we obtain the identity of that image on the disk that best matches the input. It is important to distinguish these two separate phases of

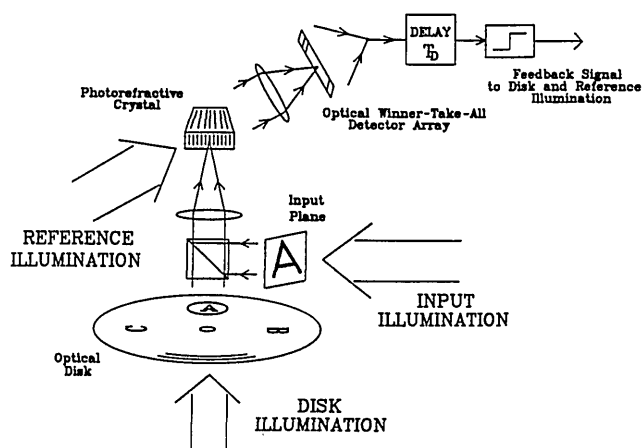
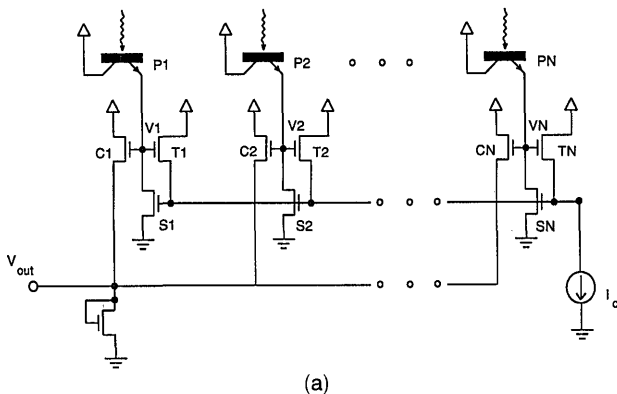


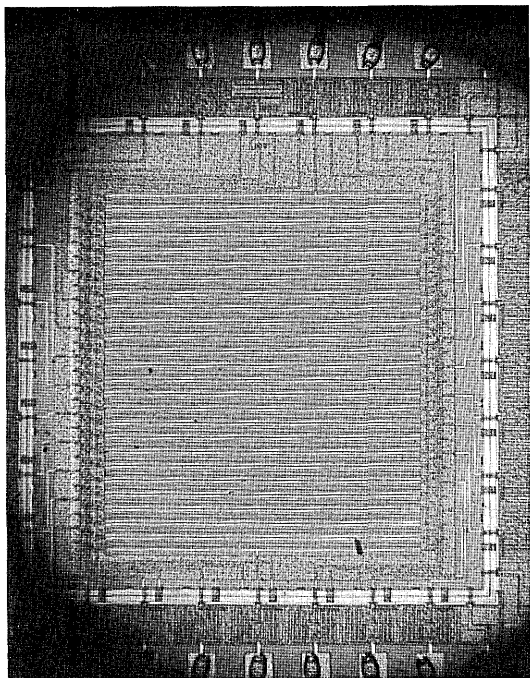
Fig. 9. Schematic of optical-disk-based image associative memory.

data reduction. The first phase takes place in the output detector array and is basically a winner-take-all (WTA) operation over each column of correlation data, while the second phase takes place in the feedback loop and comprises a WTA operation over one disk rotation. The first phase is accomplished by using a WTA detector array that is fabricated on a custom analog VLSI chip and is described in detail below. The second phase is achieved by using discrete electronics in the feedback path of the system. Continuing with the system description then, we see that the output of the optical WTA detector is thresholded and delayed for one disk rotation so that, stored in the feedback electronics, is a pulse whose temporal position indicates the location on the disk of the best match with the input image. Once such a pulse is in the loop, the readout phase is completed by closing the feedback path. In this phase both the disk and reference illumination are controlled by the feedback signal so that the disk is read out once per rotation, thereby retrieving the proper stored image and at the same time reinforcing the correct hologram in the crystal. The retrieved image may be viewed on a CCD detector (not shown in the figure) that images the disk surface as it is illuminated once per disk rotation. The system output, therefore, is a flashing image of the retrieved memory.

In order to realize the required WTA function over the 1-D detector array, we use a custom analog VLSI chip. The detector array is shown in Fig. 10(a) and is simply an array of bipolar phototransistors that serve as variable current source inputs to the WTA circuit that is described by Mead.<sup>11,12</sup> In our case the current input at a given node is proportional to the intensity of the light falling on the corresponding phototransistor. Suppose phototransistor  $P_1$  receives the largest intensity. Since the gates of transistors  $S_1-S_N$  are all held at the same potential  $V_c$ , node  $V_1$  must increase its voltage in order that  $S_1$  may support the additional current from  $P_1$ . When  $V_1$  increases,  $I_1$  must also increase. Since transistors  $T_1-T_N$  must all share the total current  $I_c$ , a competition is established. As the currents  $I_2-I_N$  decrease in response to the demands of  $I_1$ , the voltages  $V_2-V_N$  must also decrease, which establishes the feedback mechanism through which node  $V_1$  eventually becomes the only nonzero voltage among  $V_1-V_N$ . Transistors  $C_1-C_N$  serve to convert the voltages  $V_1-V_N$  to currents, which are then summed on the common line  $I_{sum}$ . In steady state, this sum current is entirely owing to the winning voltage that can be detected at node  $V_{out}$ . Owing to the subthreshold operation of the circuit, the output of the WTA array is a voltage proportional to the logarithm of the largest detected intensity. The reader is referred to Ref. 12 for a more detailed description of the operation of the circuit shown in Fig. 10(a). In Fig. 10(b) we show a photograph of the WTA detector array; the array was fabricated by using the metal-oxide semiconductor implementation service (MOSIS) facility. The array shown comprises 25 detector elements, each of which can be seen as a long rectangular region



(a)



(b)

Fig. 10. Winner-take-all detector array: (a) circuit diagram, (b) photograph of chip.

in the figure. Each detector was designed to be approximately 1 mm in length in order to minimize lateral misalignment errors. Upon testing, the WTA output of the array exhibited a response time of 80  $\mu$ s in response to a change in the largest detected intensity of 40  $\mu$ W and exhibited good linearity. Following the WTA detector array, the second phase of data reduction takes place. This phase involves thresholding the WTA output so that only a single pulse is generated during each disk rotation. Although our experimental demonstrations utilized a fixed threshold, an adaptive threshold should be used in practice so that, independent of the absolute signal levels, at least one pulse is present in the feedback loop at all times. Using such an adaptive threshold guarantees that the system retrieves an association in exactly one disk rotation. Furthermore, the adaptive threshold can be used to support the retrieval of multiple associations that will be temporally multiplexed at the system output.

## B. Hologram Erasure During the Compare Phase

During the second phase of the associative memory (AM) operation described above, the disk illumination that reads out the hologram stored in the PR crystal begins to erase this stored hologram. Since this hologram comprises an array of gratings that correspond to only one image on the disk, cross correlation with the other elements of the reference library constitutes an incoherent hologram erasure mechanism. It is true that coherent erasure takes place during that portion of a disk rotation corresponding to the autocorrelation time; however, this time is small compared with the total disk-rotation time. These observations permit us to set an upper bound on the readout intensity that can be used during this phase. If an intensity greater than this bound is used, then correlation values measured near the end of a disk rotation are diminished as compared with those values measured near the beginning owing to the loss of hologram diffraction efficiency incurred during this period of illumination. In order to estimate this intensity bound we observe that the only measurable change in the correlation signal is one that is larger than the system noise floor. This implies that

$$\eta(T_c)I_c \geq \eta(0)I_c - T(\text{SNR}), \quad (20)$$

where  $\eta(t)$  is the effective hologram diffraction efficiency at time  $t$ ,  $T_c$  is the disk-rotation time,  $I_c$  is the readout intensity,  $T(\text{SNR})$  is the smallest resolvable change in intensity measurable at the WTA detector array, and SNR is the signal-to-noise ratio. Since for incoherent erasure we have  $\eta(t) = \eta(0)\exp(-\beta I_c t)$ , where  $\beta$  is a decay constant and is measured to be  $\beta = 0.2 \text{ (mm}^2/\text{mW)}/\text{s}$  in our system, we can arrive at an expression for the maximum  $I_c$  as

$$I_c = \frac{T(\text{SNR})}{\eta(0)[1 - \exp(-\beta T_c I_c)]}. \quad (21)$$

This transcendental equation may be solved graphically for the values in our system to yield  $I_c = 0.6 \text{ mW}/\text{mm}^2$ . This readout intensity is expected to give rise to a correlation peak power of 11.9  $\mu$ W, so we see that the minimum hologram erasure constraint does not seriously affect the detectability of the correlation peak.

## C. Experimental Demonstration of Closed-Loop Optical-Disk-Based Associative Memory

The operation of the optical AM described in Subsection 3.A was demonstrated by using the PR correlator of Section 2 together with discrete feedback electronics to close the loop. During these closed-loop demonstrations, the WTA detector array had not yet been fabricated, so a single Photodyne 1500XP photodetector was used instead. This substitution results in a loss of shift invariance in the vertical direction (across-track) in our demonstration system. Use of the appropriate detector array recovers this shift invariance. The reference library for these experiments is



shown in Fig. 11 and comprises the eight memory images shown. The disk in this case was constructed from a Kodak 4 in.  $\times$  5 in. (10.16 cm  $\times$  12.7 cm) photographic plate exposed with the desired memories in order to achieve transmissive operation. The effective disk radius for this experiment therefore was 5 cm, and the reference images measured approximately 1.5 cm  $\times$  1.5 cm each. The two memories indicated in the figure as memory 1 and memory 2 were chosen for their large degree of similarity. In order to characterize the AM system operation these two memories are used.

Shown in Figure 12 are oscilloscope traces representing one disk rotation worth of data obtained during the compare phase of the AM system operation. In this figure the correlation signal for each memory is produced sequentially with memory 1 occurring approximately 26 ms before memory 2. Figure 12(a) is the compare-phase data obtained when memory 1 is used as an input to the system, and Fig. 12(b) is the same for memory 2. Notice the large degree of cross correlation indicated by the triangular signals in the figures. This is what we expect given the similarity in the two memories; however, also notice the strong autocorrelation peaks recorded. The presence of these strong peaks indicates that a simple fixed threshold can be used to squelch the cross-correlation data and to provide a single pulse in the feedback loop. This feedback pulse was seen to cause the AM loop to successfully lock onto the correct stored memory in all cases. Since the compare phase of this system lasts exactly one disk-rotation time, using a disk-rotation rate of 10 Hz provided a convergence time for our demonstrations of 0.1 s. We should note here that horizontal shift invariance (along-track) is provided in the system and can be imagined as resulting in a small time delay added to the correlation signal. Large time delays therefore represent different memories, while small delays are used to retrieve shifted versions of the stored images.

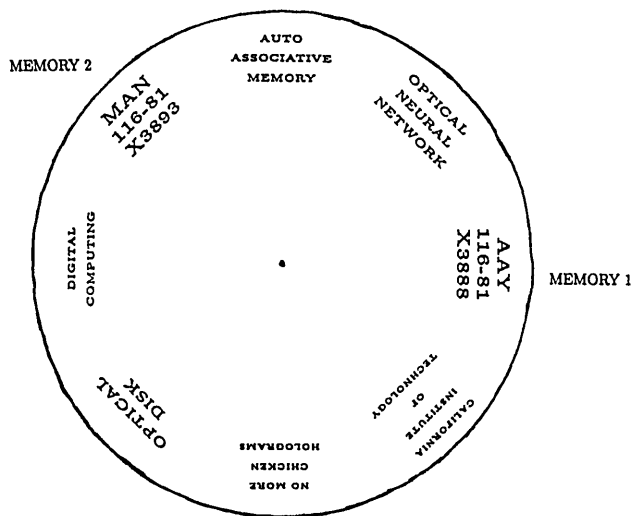
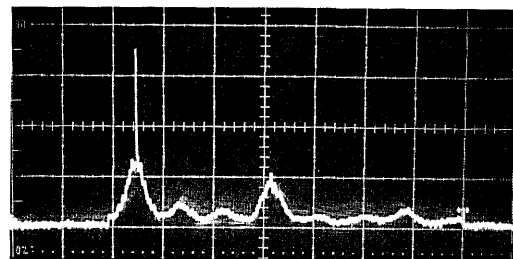
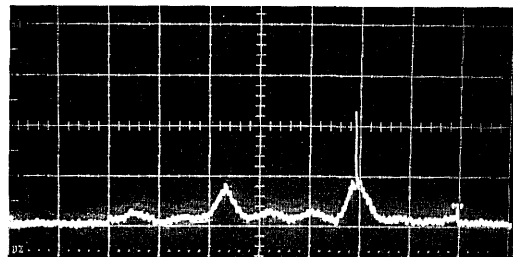


Fig. 11. Reference image library used in the associative memory experiments.



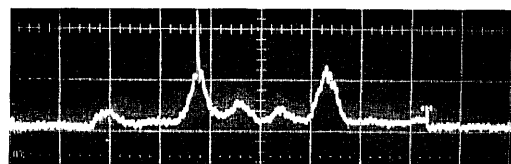
(a)



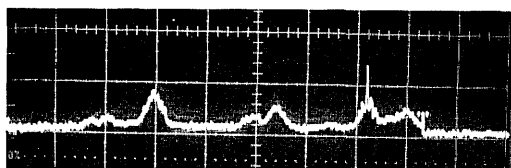
(b)

Fig. 12. Compare-phase data taken from AM: (a) memory 1 as input, (b) memory 2 as input.

An attractive feature of any AM is its ability to retrieve an appropriate stored memory from a partial input image. Shown in Fig. 13 are the correlation signals associated with partial inputs from memory 1. Figure 13(a) corresponds to the first two lines of memory 1 acting as a partial input, and Fig. 13(b) corresponds to just the first line of memory 1 presented as a partial input to the AM. In Figs. 14(a) and 14(c) we show the partial inputs from memory 1 and memory 2, respectively, along with the output state of the locked AM system in response to each in Figs. 14(b) and 14(d). It is interesting to note that, when the partial input to the AM system is line number 2 of memory 1, both memory 1 and 2 are retrieved. When both memories are locked in the loop, the system is seen to flash twice per rotation, alternately illuminating memory 1 and 2. The output CCD therefore detects the two retrieved images flashing in sequence. This is of course the behavior we expect and indicates the capability of the system

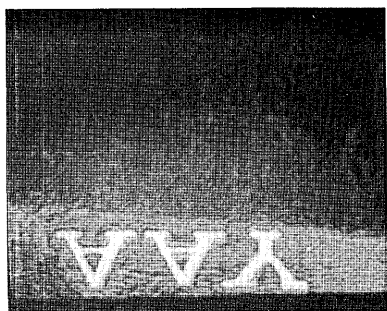


(a)

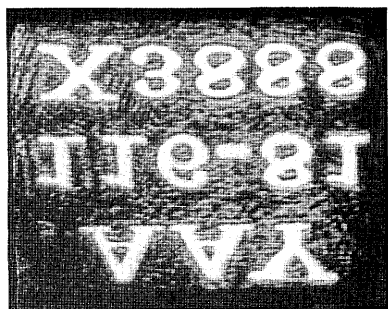


(b)

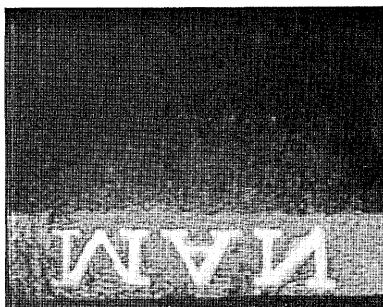
Fig. 13. Compare-phase data taken from AM with partial inputs: (a) first two lines of memory 1, (b) first line only of memory 1 as input.



(a)



(b)



(c)



(d)

Fig. 14. Image completion with the disk-based AM: (a) partial input of memory 1, (b) completed image retrieved from the system; (c) partial input of memory 2, (d) completed image retrieved from the system.

for retrieving multiple associations (multiplexed in time) when presented with an ambiguous input image.

#### 4. Effects of Disk Imperfections

The optical-disk systems that are commercially available are designed to specifications consistent with a serial readout application. Error correction coding

is extensively used to deal with media defects, and feedback is incorporated to offset mechanical imperfections such as wobble, center offset, and coating thickness variations in real time. These same error sources and imperfections can affect the performance of architectures we have been discussing. Although revised design specifications for commercial systems would eliminate the need for much of the following discussion, the Sony system used in this research does suffer from the above-mentioned limitations; therefore a discussion of the effect of such imperfections is included.

In Ref. 1 we characterized the optical disk in terms of the uniformity of its coating. We found that for glass-covered Sony disks, thickness variations were negligible over areas of several millimeters squared. Since areas of this size may contain up to a million pixels, we conclude that coating thickness variations do not play a role in determining the behavior of the systems discussed here. The thickness variations measured in Ref. 1 could be reduced even further by more careful processing during manufacturing, thereby improving the phase characteristics of the optical-disk SLM. In contrast to the lack of care taken to ensure good coating thickness uniformity in commercial optical-disk systems, media defects are minimized in conventional disks. Since the storage industry is concerned with long-term high-density memory with high integrity, media defects are minimized in commercial systems, and sophisticated error checking and correction is employed to reduce the system error rate even further. A typical level of raw media imperfections in a commercial system is  $10^{-4}$  bit error rate and can be compensated by using a variety of error correcting codes to achieve a bit error rate of  $<10^{-12}$  at the system level. For the correlation architectures discussed here, media defects may be treated as a noise source in the reference plane and as such are already compensated to some extent by virtue of the recall mechanism itself. It is well known that the correlation filter, or matched filter, is the optimal decoder for a channel that is corrupted by stationary noise. The inherent readout mechanism employed when one uses an optical disk in an image correlation system therefore is optimal for the tolerance of media defects when such defects may be characterized as arising from a stationary noise source.

The previous discussion regarding the effects of disk coating imperfections and media defects has permitted us to reduce the number of nonnegligible disk imperfections to just two: disk wobble and disk center offset. Disk wobble and disk center offset are mechanical imperfections in the disk planarity and in track centering, respectively. For this application, disk center offset is of no consequence since the correlation function is invariant to reference-image shift, but disk wobble is critical. We may consider disk wobble as resulting from a warping of the disk surface. It can also result from mechanical imperfections in the rotation mechanism; however, we combine these effects and discuss them together.

Wobble can be characterized by the maximum magnitude and maximum frequency of  $\Omega(\theta)$ , where  $\Omega(\theta)$  is the disk surface inclination angle (i.e., the angle between a radial line of the disk and a horizontal plane) as a function of disk-rotation angle. For the Sony system,  $\max[\Omega(\theta)] = 1^\circ$  and the maximum frequency is 50 cycles/rotation. Significant alignment error comes about by virtue of wobble-induced disk plane tilt. Disk tilt results in a shift of the reference Fourier transform. As an aside we should note that this Fourier plane shift does take place if a transmissive disk is used. In the case of a reflective disk-based system, however, this shift does cause a misalignment between the readout illumination and the stored hologram. The misalignment is easily quantified as  $\Delta x' = F \sin \theta$ , where  $\theta$  measures the disk tilt angle ( $< 1^\circ$  for Sony media) and  $x'$  measures the shift of the Fourier transform in the crystal plane. We may also calculate the Fourier-plane alignment sensitivity as the effective resolution spot size in the crystal. This is given by  $\Delta x'_{\text{MAX}} = 1.22\lambda F/A$ . We may determine a limit on permissible disk tilt by setting  $\Delta x' < \Delta x'_{\text{MAX}}$ , and we find that  $\sin \theta < 1.22\lambda/A$ . So we see that wobble tolerance can be traded off for disk SBP usage. Since wobble tolerance requires a small Fourier-transform system aperture  $A$ , this in turn implies a low system cutoff frequency that results in a minimum disk-plane resolution of  $\Delta x > 1.22\lambda F/A$ . For typical system parameters, we have  $\theta = 1^\circ$ , yielding  $A < 44 \mu\text{m}$ , which is smaller than a  $100 \times 100$  pixel image on the disk. Another consequence of enlarging the spot size in the Fourier plane to tolerate disk tilt is the concomitant reduction in writing energy and the resulting longer exposure times.

We show a candidate wobble-compensation scheme in Fig. 15. Although the technology already exists for performing the requisite wobble correction, this simple exercise clarifies the engineering requirement of such a system. If we insist that  $F/A \approx 1$  to ensure good disk SBP usage, then we have  $\Delta x < 1 \mu\text{m}$ . The system of Fig. 15 can achieve real-time alignment compensation in the Fourier plane by adjusting the disk illumination angle according to the instantaneous disk tilt as measured by the probe beam shown. The probe beam may be incident from the same side as the readout illumination as well; however, it is shown probing the opposite side of the disk for clarity. If the piezo mirror rotates by an amount equal to the disk tilt, then the reflected light will always be aligned with the  $z$  axis as shown. Piezo mirrors with the capability of providing up to  $1^\circ$  of angular displacement and with response times fast enough to respond to the probe signal ( $< 200$  Hz) are readily available. The requirement for  $1\text{-}\mu\text{m}$  positioning accuracy in the Fourier plane and 1000 resolvable spots on the disk translates to a piezo-mirror angular accuracy of  $\theta_{\text{min}} \approx 0.7$  mrad. The probe beam therefore is required to detect a disk tilt of 0.7 mrad out of  $1^\circ$ , which requires a detector sensitivity of at least 1 in 26.

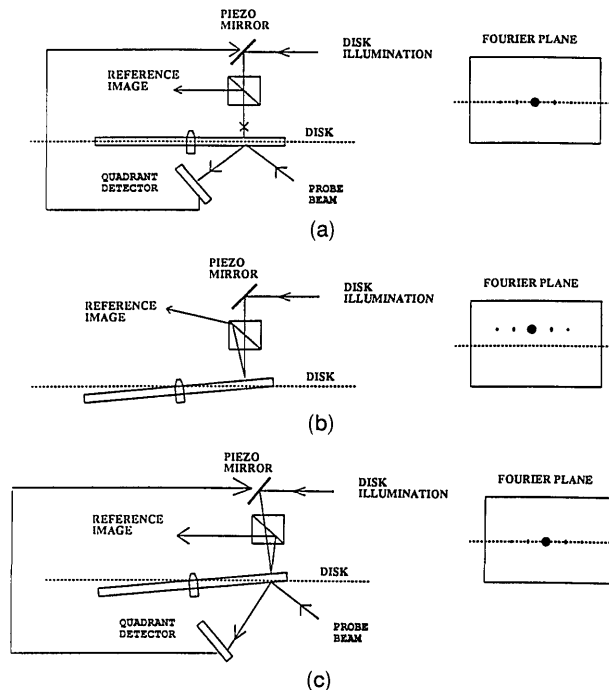


Fig. 15. Wobble-compensation scheme for the PR correlator: (a) no disk wobble, (b) disk wobble causing a shift in the reference image spectrum, (c) piezo mirror compensating for disk wobble.

## 5. Conclusions

In this paper we have presented an optical-disk-based image correlator capable of providing very high correlation rates (400,000 images/s) with good light efficiency. On the basis of this image correlator we have described an optical-disk-based associative memory that takes advantage of the parallel-access capabilities afforded by optical storage media. The capacity of this associative memory is given by the capacity of the optical-disk image library. The retrieval time for recovery of a single image association is given by the disk-rotation time and is approximately 10 ms for conventional disk drives. One attraction of the present system is its full 2-D shift invariance. The Fourier-transform hologram in the crystal provides vertical shift invariance, while disk rotation provides shift invariance in the horizontal dimension. Since vertical position of the input image can be detected as the location of the winner in the optical WTA detector, this information can be stored for later retrieval. Horizontal position is represented as a temporal delay, which in turn determines when the closed-loop disk illumination is pulsed. In this way, a shifted version of one of the stored images can be retrieved from the system. Another attractive feature of this system is the locking or stable nature of the closed-loop operation. Since both the disk illumination and the reference beam are pulsed together, the hologram stored in the crystal is reinforced during each disk rotation. This ensures that the readout diffraction efficiency remains stable and guarantees that correlation-plane SNR does not degrade. This locking has a further advantage. If an incomplete or corrupted

input is presented to the system, the loop locks to a complete, uncorrupted association, which in turn reinforces the hologram in the crystal. In this way, the hologram of the distorted input is slowly corrected through the locking action of the loop.

This work was funded by the U.S. Army Research Office. We thank R. R. Neurgaonkar for the SBN crystal used in the experiments. The authors also thank Alan Yamamura, Seiji Kobayashi, and the Sony Corporation for assistance with the optical-disk-writing system.

## References

1. D. Psaltis, M. A. Neifeld, A. Yamamura, and S. Kobayashi, "Optical memory disks in optical information processing," *Appl. Opt.* **29**, 2038–2057 (1990).
2. D. Psaltis, M. A. Neifeld, and A. Yamamura, "Image correlators using optical memory disks," *Opt. Lett.* **14**, 429–431 (1989).
3. A. A. Yamamura, M. A. Neifeld, S. Kobayashi, and D. Psaltis, "Optical disk based artificial neural systems," *Opt. Comput. Process.* **1**, 3–12 (1991).
4. D. Gabor, "Associative holographic memories," *IBM J. Res. Dev.* **13**, 156–159 (1969).
5. Y. Owechko, G. Dunning, E. Marom, and B. Sofer, "Holographic associative memory with nonlinearities in the correlation domain," *Appl. Opt.* **26**, 1900–1910 (1987).
6. K. Hsu, H. Li, and D. Psaltis, "Holographic implementation of a fully connected neural network," *Proc. IEEE* **78**, 1637–1645 (1990).
7. D. Anderson, "Coherent optical eigenstate memory," *Opt. Lett.* **11**, 56–58 (1986).
8. A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory* **IT-10**, 139 (1964).
9. J. Yu, F. Mok, and D. Psaltis, "Capacity of optical correlators," in *Spatial Light Modulators and Applications II*, U. Efron, ed., *Proc. Soc. Photo-Opt. Instrum. Eng.* **825**, 114–120 (1987).
10. C. Gu, J. Hong, and S. Campbell, "2-D shift invariant volume holographic correlator," *Opt. Commun.* **88**, 4646–4648 (1992).
11. J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of  $O(N)$  complexity," in *Advances in Neural Information Processing I*, D. S. Touretzky, ed. (Kaufmann, San Mateo, Calif., 1989), pp. 703–711.
12. C. Mead, *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, Mass. 1989), pp. 67–78.