

Optical memory disks in optical information processing

Demetri Psaltis, Mark A. Neifeld, Alan Yamamura, and Seiji Kobayashi

We describe the use of optical memory disks as elements in optical information processing architectures. The optical disk is an optical memory device with a storage capacity approaching 10^{10} bits which is naturally suited to parallel access. We discuss optical disk characteristics which are important in optical computing systems such as contrast, diffraction efficiency, and phase uniformity. We describe techniques for holographic storage on optical disks and present reconstructions of several types of computer-generated holograms. Various optical information processing architectures are described for applications such as database retrieval, neural network implementation, and image correlation. Selected systems are experimentally demonstrated. *Key words:* Optical memory disk, spatial light modulator, optical computing, computer-generated holograms.

I. Introduction

Ever since the introduction of the videodisc system in the late 1970s and the compact audio disk player in the mid-1980s, optical disk technology has been maturing at a rapid pace. Both write-once read-many (WORM) and magneto-optic read/write disk drives are presently available for high density storage on mainframes and personal computers. The conventional mode of both reading and writing used in present optical disk systems is serial. Specifically, a laser source will write one bit of data at a time on the disk, typically through a thermal mechanism.¹ Readout is achieved by using a lower power beam to illuminate the location of each bit on the disk individually and, based on the reflected or transmitted intensity detected, the bit is decoded as a logical 1 or 0. Although serial readout is well suited to conventional computers, the optical disk itself is naturally a parallel readout device.^{2,3} To see this, consider illuminating a large portion of the disk with a collimated beam. The reflected or transmitted light contains all the data originally recorded in the illuminated area and a simple imaging system makes these data available to a detector array. This parallel access capability can be attractive when trying to solve memory access and contention problems in parallel computing architectures or when try-

ing to implement an intelligent memory search procedure as with database machines.⁴ Further, the optical disk represents a high resolution, computer controllable, spatial light modulator (SLM) which may be used in various optical computing architectures. For example, images stored on an optical disk may serve as a library of references in an optical image correlator and holograms stored on the disk may serve as interconnect patterns for hybrid optical/VLSI based neural networks.

In this paper we discuss the application of optical disk technology to areas in which parallel retrieval may be advantageous. We begin by characterizing the disk system used in our work, a Sony prototype sampled format drive with both WORM and magneto-optic media. In Sec. III we discuss parallel optical readout of 2-D blocks of data such as images. In the same section, we describe the use of optical disks as holographic storage media. We present and analyze several techniques for storing and retrieving data holographically and suggest some applications of holographic disk based storage. In Sec. IV we describe the use of optical disks as both storage and interconnect elements in neural network architectures. Finally, optical disk based image correlators are described and demonstrated in Sec. V. All the applications we discuss here are designed to combine the parallelism and interconnectivity of optics with a mature optical disk technology to result in feasible optical systems that perform useful computational tasks.

II. Characterization

The prototype Sony disk system used in most of our work (Fig. 1) can read and write both write-once and magneto-optic 5-in. reflective optical disks. The system records data as circular $1\text{-}\mu\text{m}$ diam pixels along a

Seiji Kobayashi is with Sony Corporation, Tokyo 100-31, Japan; the other authors are with the California Institute of Technology, Pasadena, California 91125.

Received 17 July 1989.

0003-6935/90/142038-20\$02.00/0.

© 1990 Optical Society of America.

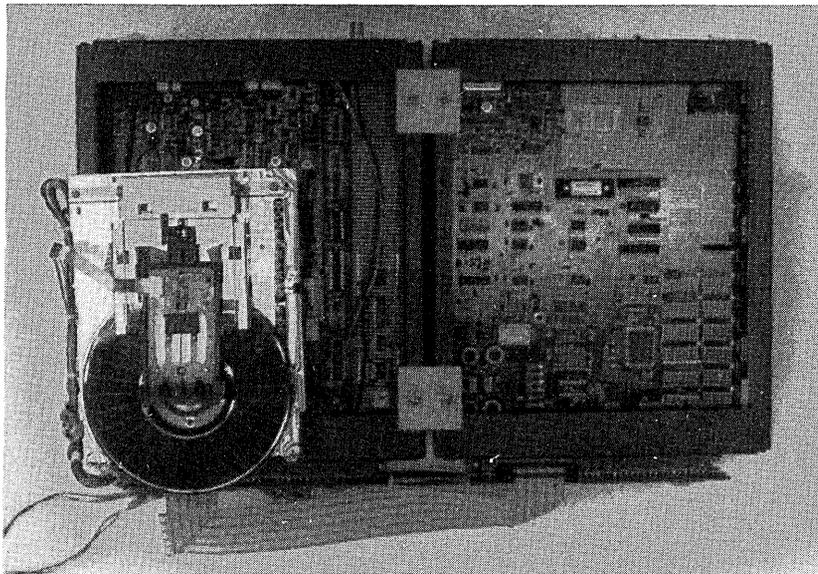


Fig. 1. Sony prototype optical disk system.

spiral on the disk with 20,000 turns between a 3-cm inner radius and 6-cm outer radius. The disk is divided into thirty-two sectors, and each loop of the spiral from the beginning of sector 0 to the end of sector 31 is called a track. Because the radius of the tracks changes gradually with angle, we often model the tracks as concentric circles separated by $1.5 \mu\text{m}$. Pixels are recorded with a constant angular separation of $.001^\circ$. This corresponds to an along track pixel-to-pixel separation that varies between $.5 \mu\text{m}$ and $1 \mu\text{m}$ depending on radial position on the disk. This pixel recording density yields a storage capacity of over 7×10^9 bits on each side of the disk. The system is interfaced to a personal computer (PC) which provides serial read/write access to the disks. The system can read or write up to 15 million bits/s. Consequently a 1000×1000 image can be entered on the disk in $1/15$ s. Since we can only write one line of the image per revolution, about 30 s are required to record the image in 2-D format on the disk. Note, however, that a thousand images using the same tracks could also be written during the same amount of time.

A variety of materials and recording mechanisms have been proposed for use in optical disks.¹ We briefly describe the recording mechanism employed in the write-once disk that we use in our experiments (see Ref. 5 for further details). The disk contains four thin metal alloy films of Sb_2Se_3 (300 Å), Bi_2Te_3 (150 Å), Sb_2Se_3 (1400 Å), and Al (1000 Å) formed by sputtering deposition on a glass or plastic substrate. The thickness of the various layers is chosen so that they form a low (5%) reflectivity interference filter. During the recording stage, a focused laser beam heats a spot of the Bi_2Te_3 layer through absorption. The Bi_2Te_3 and Sb_2Se_3 then form a four-element alloy by diffusion, eliminating the sharp interfaces between the layers. The low reflectivity interference filter is thus destroyed increasing the reflectivity of the medium to 12%. This reflectivity difference is detected during

readout and decoded as a logical 1 or 0. The reflectivity of an interference filter is wavelength dependent. Our quoted figures are for the 633 nm He-Ne illumination that we use, but the thicknesses of the layers are chosen to maximize the change in reflectivity for the laser diode wavelength of 830 nm.

The magneto-optic disk contains a rare earth transition-metal alloy of TbFeCo. During the recording stage, the laser heats a spot on the disk above 180°C , the Curie temperature of the material. As the spot cools below this temperature, the material within the spot retains the magnetization of an external field applied perpendicular to the disk surface. The polarization of a low power readout laser rotates on reflection from the spot by an angle of $\pm 0.15^\circ$, from the magneto-optic Kerr effect. The reflectivity of the magneto-optic disk is 17%, and the sign of the rotation angle depends on the direction of magnetization in the spot. This rotation is detected through a crossed polarizer and decoded as a logical 1 or 0. Depending on the setting of the polarizers, the amplitude of the light corresponding to the two states can be either on/off or plus/minus.⁶

Current disk systems use either continuous or sampled format schemes to maintain the position of the head over data in a track. Continuous format systems use a return signal either from a guide-groove embossed on the disk or the recorded data itself to constantly monitor and correct the position of the head relative to the data in a track. In contrast, sampled format systems, such as our experimental one, use tracking and timing information embossed along radial lines on the disk to periodically monitor and correct the head position. These lines of tracking and timing information appear every 270 pixels. Each line consists of a pattern of three embossed pits repeated in all 20,000 tracks as shown in Fig. 2. The first two pits provide tracking information. They are displaced an equal distance from the center of the track, one toward

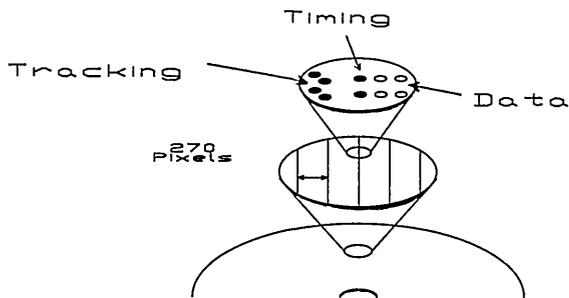
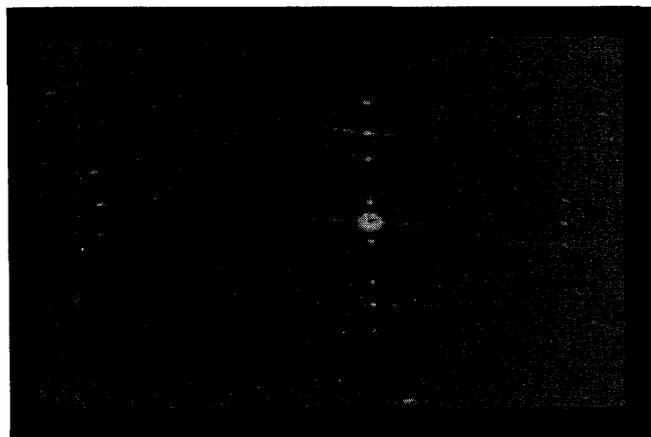


Fig. 2. Sampled-format tracking and timing information system.

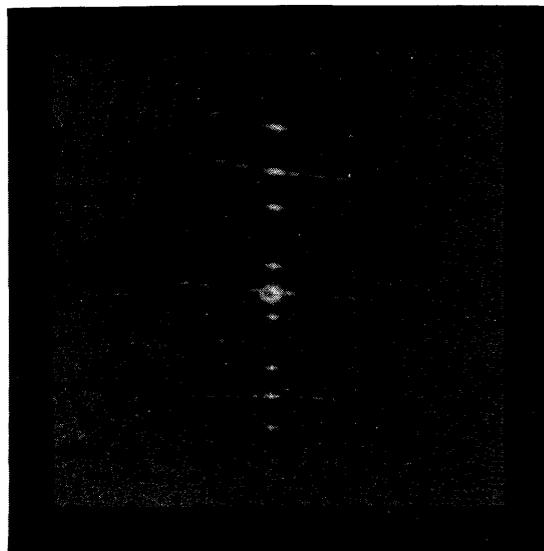
the inside of the disk and the other toward the outside. If the head is exactly over the center of the track, the readout signal strength of the two pits will be equal; otherwise, the signal returning from one of the pits will be stronger than the other, thus indicating the direction to move the head. The third pit provides timing information used to synchronize the system clock and the disk.

There are two byproducts of the sampled format scheme that facilitate the parallel readout of data. First, the across track alignment of tracking and timing information combined with the synchronization between recorder and disk rotation allows us to specify the position of individual pixels with submicron accuracy in any one of more than one billion locations. This provides us with across track coherence, the ability to radially align pixels across different tracks. In contrast, with continuous format systems, the position of pixels in different tracks can drift by several pixel widths within a single sector. Second, the absence of guide-grooves allows us to retrieve high contrast images from low contrast media through schlieren imaging as described in Sec. III.

When we consider using the disk as a spatial light modulator, a number of additional performance issues arise. The resolution is determined by the track spacing in the radial direction ($1.5 \mu\text{m}$) and the minimum spot size in the azimuthal direction ($0.5\text{--}1 \mu\text{m}$). Notice that there is an inherent sampling in the radial direction due to the tracks. At the outer tracks, where the recorded pixels do not overlap, the image is also sampled in the azimuthal direction. We see later that we can make use of image diffraction caused by this sampling. Figure 3 shows the far field diffraction pattern when a grating recorded on the disk is illuminated. The grating was formed by periodically recording two tracks with all pixels on followed by two tracks with all pixels off. In Fig. 3(a), the grating is recorded on the inner tracks where pixels overlap along the track. In this case, the image is sampled in only one dimension thus producing diffraction orders in one dimension only. Figure 3(b) shows the diffraction with the same grating recorded at the outer edge of the disk. In this case, sampling in both dimensions results in a 2-D diffraction pattern. The maximum spatial frequency that can be recorded without aliasing of the image spectra is one-half of the sampling frequency in each direction.



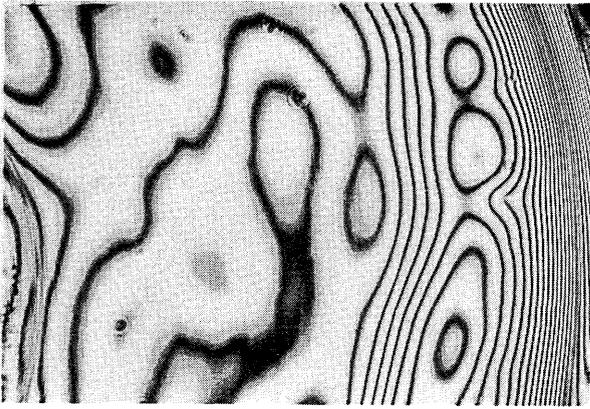
(a)



(b)

Fig. 3. Far field diffraction pattern from grating: (a) grating written on inner tracks ($R = 3 \text{ cm}$); and (b) grating written on outer tracks ($R = 6 \text{ cm}$).

The reflectance function of the disk is basically binary both for the write-once and the magneto-optic disks. We have observed some dynamic range in the reflectivity of the write-once disks, controllable by varying the exposure for each pixel. We have not yet characterized fully the grey scale capability of the system. In any case, some form of area modulation can be used to encode multiple grey levels at the expense of space-bandwidth product. We will demonstrate one such method in the following section. The contrast of the light reflected from the disk is low for the Sony write-once disks (2:1). For magneto-optic disks, the polarization of the modulated light is orthogonal to the polarization of the incident light and the use of orthogonal polarizers in conjunction with the carrier encoding method discussed in the next section yields excellent contrast, limited primarily by the quality of the polarizers.



5mm

Fig. 4. Interferogram of Sony WORM disk.

Many of the processing architectures we propose use coherent processing techniques requiring phase uniformity across the surface of the disk. We have used a Fizeau interferometer to measure the phase uniformity of the Sony disks. Figure 4 shows a Fizeau interferogram of a 24- × 36-mm area of a glass-covered Sony write-once disk taken with a He-Ne laser source providing illumination. The figure shows numerous regions on the disk with optical thickness variations of less than a wavelength ($\lambda = 633$ nm) over distances of several millimeters. These regions are sufficiently large to contain images or holograms thousands of pixels on a side. The greater phase variation towards the outer edge of the disk is most likely caused by index variations due to stresses induced during manufacturing. We have also tested plastic-covered disks which generally show greater phase variation than the glass-covered ones.

In most uses, it would be more convenient if the optical disk system recorded pixels on a Cartesian grid. However, as noted earlier, our system actually writes pixels along curved tracks. We can neglect this curvature if we restrict attention to a small area of the disk. Consider a region at a distance R from the disk center. As shown in Fig. 5, we establish Cartesian coordinate axes with x in the azimuthal or along track direction and y in the radial or across track direction. Equation (1) converts the polar coordinates of the disk to the Cartesian coordinates in the region of interest:

$$\begin{cases} x = r \sin(\theta) \\ y = r \cos(\theta) - R \end{cases} \quad (1)$$

The center-to-center spacing of the pixels in the radial dimension is δ_r , and the angular separation between adjacent pixels is δ_θ in azimuth. We now superimpose a Cartesian grid on this pixel structure with x and y spacings as follows:

$$\begin{cases} \Delta_x = R\delta_\theta \\ \Delta_y = \delta_r \end{cases} \quad (2)$$

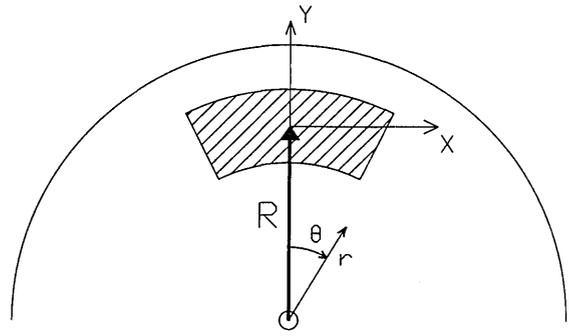


Fig. 5. Coordinate system for calculating the effect of track curvature.

This choice for Δ_x and Δ_y provides the best match between the pixels recorded on the disk and the points on the Cartesian grid. We now calculate the deviation of the pixel locations from their presumed Cartesian locations. The presumed coordinates of the points on the Cartesian grid are

$$\begin{cases} x' = n\Delta_x \\ y' = m\Delta_y \end{cases} \quad (3)$$

whereas the actual location of the pixels on the polar grid are

$$\begin{cases} r = R + m\delta_r + n \frac{\delta_\theta \delta_r}{2\pi} \\ \theta = n\delta_\theta \end{cases} \quad (4)$$

The actual Cartesian coordinates of the recorded pixels, therefore, are

$$\begin{cases} x = \left(R + m\delta_r + n \frac{\delta_\theta \delta_r}{2\pi} \right) \sin(n\delta_\theta) \approx n\Delta_x + nm \frac{\Delta_x \Delta_y}{R} \\ y = \left(R + m\delta_r + n \frac{\delta_\theta \delta_r}{2\pi} \right) \cos(n\delta_\theta) - R \approx m\Delta_y + n \frac{\Delta_x \Delta_y}{2\pi R} - n^2 \frac{\Delta_x^2}{2R} \end{cases} \quad (5)$$

We calculate the deviation between the actual pixel position and the presumed location on the Cartesian grid by subtracting Eq. (5) from Eq. (3):

$$\begin{cases} \epsilon_x = x' - x \approx nm \frac{\Delta_x \Delta_y}{R} = \frac{x'y'}{R} \\ \epsilon_y = y' - y \approx n \frac{\Delta_x \Delta_y}{2\pi R} - n^2 \frac{\Delta_x^2}{2R} = x' \frac{\Delta_y}{2\pi R} - \frac{x'^2}{2R} \end{cases} \quad (6)$$

For an array of 1000×1000 pixels on the Sony disks, the worst case pixel placement error is 1.25% of the array size (12.5 pixels) in the x -direction (at $R = 3$ cm, $\Delta_x = 0.5 \mu\text{m}$, and $\Delta_y = 1.5 \mu\text{m}$) and 0.14% of the array size (1.4 pixels) in the y -direction (at $R = 6$ cm, $\Delta_x = 1 \mu\text{m}$, and $\Delta_y = 1.5 \mu\text{m}$). In applications where this kind of positional error is not tolerable, we need to compensate the curvature through optical techniques and/or the recording geometry. We describe such methods in greater detail in the following section.

Diffraction efficiency is a key parameter in determining overall system efficiency since many of the

optical systems presented in the following sections use light diffracted from the disk. Given an accurate model of the surface reflectivity of the disk and how it will be used in an optical system, we can calculate the expected efficiency of the disk in that application. We model the reflectivity pattern of the disk using the following equation:

$$r(x,y) = r_0 + (r_1 - r_0) \left[\sum_{n,m} b_{nm} \delta(x - n\Delta_x, y - m\Delta_y) \right] \otimes s(x,y), \quad (7)$$

where b_{nm} represents the binary pixel pattern, r_0 and r_1 the reflectivity of unwritten and written pixels, respectively, \otimes the convolution operation, and $s(x,y)$ the shape of each pixel.

The light reflected from the surface of the disk $E_r(x,y)$ is the product of the reflectivity pattern of Eq. (7) with the field of the illuminating light beam $E_i(x,y)$:

$$E_r(x,y) = E_i(x,y)r(x,y). \quad (8)$$

We can use Fresnel diffraction to calculate the field due to light reflected from the disk at any distance from the disk, as follows:

$$E(x,y,z) = \iint E_r(x',y') \frac{z \exp(jkl)}{j\lambda l^2} dx'dy', \quad (9)$$

$$l^2 = (x - x')^2 + (y - y')^2 + z^2. \quad (10)$$

The efficiency η of the disk can then be found by integrating the intensity of light reflected by the disk over the region of space Σ where the optical system captures reflected light and dividing by the incident light energy:

$$\eta = \frac{\iiint_{\Sigma} |E(x,y,z)|^2 dx dy dz}{\iint |E_i(x,y)|^2 dx dy}. \quad (11)$$

In the Appendix, as an example, we estimate the diffraction efficiency of the disk for schlieren imaging of the first diffracted order. Substituting parameters for the Sony write-once disk into Eq. (A7), [$|r_1|^2 = 0.12$, $|r_0|^2 = 0.05$, $\Delta_x = 0.5 \mu\text{m}$, $\Delta_y = 1.5 \mu\text{m}$, and $\Delta_r = 0.5 \mu\text{m}$ with $b(x,y) = 1$, $n = 0$, and $m = 1$], we find an estimated efficiency of $\eta_e = 0.112\%$ compared with a measured efficiency of $\eta_m = 0.114\%$. We estimate that the magneto-optic disk will be almost 1000 \times less efficient than the write-once disk in most applications. This large loss in efficiency was also observed experimentally.

III. Imaging and Holography

The fact that data can be retrieved in parallel from optical disks creates the possibility for eliminating some of the bottlenecks that currently exist in computers due to the mismatch between mass storage media and semiconductor memories.² A parallel random access memory would be one possible way to construct a parallel readout optical memory. In this case, M out of the N bits stored on a disk could be specified and retrieved simultaneously. In such a system, the apparatus that would scan the memory to realize this paral-

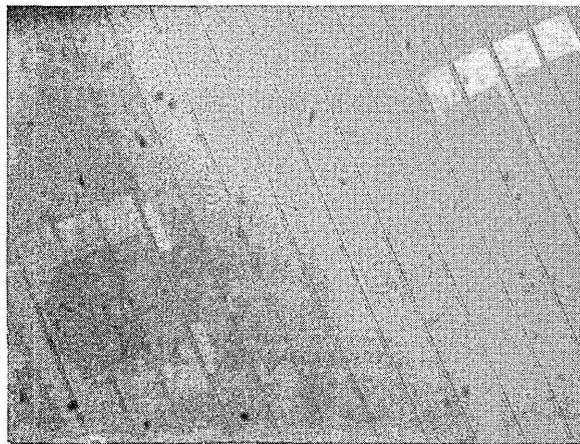


Fig. 6. Photograph of image written on Sony WORM disk.

lel retrieval capability would have to be set in $\binom{N}{M} \approx (N/M)^M$ distinct states to arbitrarily select any M -tuple. With $N = 10^{10}$ and $M = 1000$, we obtain about 10^{7000} distinct states. It is clearly not practical to realize an optical scanning mechanism that can do this. Therefore, we conclude that we must somehow structure the stored data to reduce the complexity of the access mechanism. The most straightforward way to impose such structure is to arrange the stored data in 2-D blocks, M bits each, that are retrievable in parallel. This reduces the number of choices the access mechanism addresses to a practical $\binom{N/M}{1} = 10^7$ for the previously quoted numbers. In this section, we discuss several methods, including holographic, for recording and retrieving 2-D blocks of data from optical disks.

Figure 6 is a photograph of a binary image written on the Sony write-once disk as viewed through a microscope. This image consists of 4024×512 pixels recorded on a polar grid. Note that the track curvature is not visible. The parallel lines, evident in the figure, are the radial strips of tracking and timing information described in the previous section. There are 270 pixels between each pair of these lines. Individual pixels are not discernible in this figure, but they are perfectly aligned in the radial direction resulting in the accurate recording of the letters in the figure. Note the poor contrast in Fig. 6. This is because the disk has an off-state (unwritten) reflectivity of 5%, while the on-state (written) reflectivity is only 12%. This large background and low differential reflectivity results in the poor contrast seen in the photo. Rilum and Tanguay used an interferometric technique to eliminate the background light obtained on reflection from a stamped optical disk.⁷ This technique is not applicable to the Sony disk because the recorded data do not appear as embossed pits but rather as local variations in surface reflectivity. We use an alternative means for improving the contrast of the retrieved image. Since the binary image $b(x,y)$ to be recorded on the disk is sampled by a polar grid, light reflected from the disk will be diffracted into many orders or sidebands whose center frequencies will be determined by the

grid spacing. To make this clear, consider Eq. (12) for the reflectivity of the disk:

$$r(x,y) = r_0 + (r_1 - r_0) \left[b(x,y) \sum_{n,m} \delta(x - n\Delta_x, y - m\Delta_y) \right] \otimes s(x,y), \quad (12)$$

with the x -axis in the along track direction and the y -axis in the radial direction, as before. Since the background reflectivity r_0 is not sampled, energy in the first- and high-order diffracted fields arises only from the presence of the recorded image. Therefore, an image formed by the first-order diffracted field will not contain any bias light and will have high contrast. We can calculate how much energy is diffracted into the first-order and compare this to the total incident energy to obtain an estimate for the efficiency of the disk. The Fourier transform of the reflectivity function is

$$R(u,v) = r_0\delta(u,v) + (r_1 - r_0) \times \left[B(u,v) \otimes \frac{1}{\Delta_x\Delta_y} \sum_{n,m} \delta\left(u - \frac{n}{\Delta_x}, v - \frac{m}{\Delta_y}\right) \right] S(u,v), \quad (13)$$

where $B(u,v)$ and $S(u,v)$ are the Fourier transforms of $b(x,y)$ and $s(x,y)$, respectively.

The term that contributes to the formation of the desired image is

$$R_1(u,v) = \frac{r_1 - r_0}{\Delta_x\Delta_y} B\left(u, v - \frac{1}{\Delta_y}\right) S(u,v). \quad (14)$$

From these two equations we can express the disk efficiency as

$$\eta = \iint |R_1(u,v)|^2 dudv. \quad (15)$$

For the Sony optical disk we have calculated the disk efficiency to be 0.112% (see Sec. II for parameters). This value agrees well with the measured efficiency of 0.114%. For a more detailed derivation of the efficiency of a schlieren imaging system used to image the optical disk surface, see the Appendix. An example of a high contrast image obtained by imaging the first diffracted order is shown in Fig. 7. The light diffracted by the tracks was selected to form this image.

As described in Sec. II, when we assume that pixels are written on a Cartesian grid, the presence of track curvature leads to positional errors given by Eq. (6). In the schlieren imaging system described above, the positional error of a recorded spot can lead to amplitude and phase errors in its contribution to the reflected field. We neglect the amplitude error since it only becomes significant when the position error $\bar{\epsilon}(x,y)$ is comparable to the distance between the disk and the image plane. The phase error observed in a given direction, however, can be approximated in most cases by the dot product between the wavevector in that direction \bar{k} and the position error. Since $|\bar{k}|$ is large, this phase error cannot be neglected. When we consider only the first-order diffracted beam, $k_x = 0$ and the apparent phase error over the disk is given by

$$\phi_{\text{err}}(x,y;\bar{k}) = \bar{k} \cdot \bar{\epsilon}(x,y) = -\frac{k_y x^2}{2R} = \frac{k}{2f} x^2, \quad (16)$$

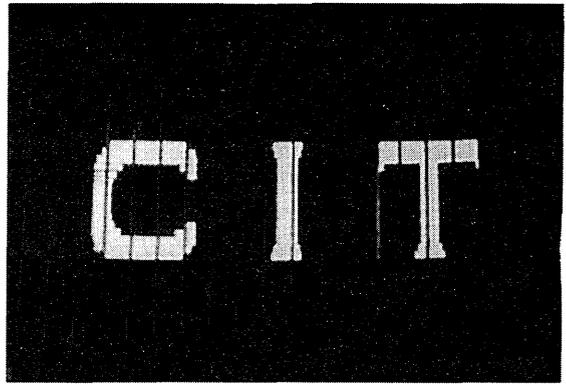


Fig. 7. High contrast image obtained by imaging the first-order diffracted component of light reflected by the disk used in Fig. 1.

$$f = -R \frac{k}{k_y} \quad (17)$$

$$k_y = \begin{cases} \frac{2\pi}{\Delta_y} & \text{for the } +1 \text{ order,} \\ -\frac{2\pi}{\Delta_y} & \text{for the } -1 \text{ order.} \end{cases} \quad (18)$$

This phase error can be modeled as a cylindrical lens at the disk plane with focal length $f = \pm R\Delta_y/\lambda$. For typical experimental parameters $R = 4.5$ cm, $\Delta_y = 1.5$ μm , and $\lambda = 633$ nm, the cylindrical focal length is 10.7 cm. This distortion can be corrected by an illumination system containing a cylindrical lens of focal length F . The product of the incident wavefront and the reflectance function of the disk in this case is

$$E_r(x,y) = E_i(x,y)\bar{r}(x,y), \quad (19)$$

where $E_r(x,y)$ is the reflected field, $E_i(x,y) = \exp(j\alpha x^2)$ is the incident field corrected by the cylindrical lens, and $\bar{r}(x,y)$ is the apparent reflectance function of the disk surface including the phase error. The illuminating optics should be chosen so that $E_r(x,y) = r(x,y)$, which yields $\alpha = -k/2f$. With this value for α , the incident illumination is given by $E_i = \exp(-jkx^2/2f)$ which can be generated by a line source located a distance f in front of the disk. A cylindrical lens with focal length F , at a distance $F + f$ in front of the disk, can be used to form the line source.

Correcting for this cylindrical lens effect does not, however, account for the positional errors of pixels due to the polar grid. In Sec. II we found that the position error of pixels in large images can exceed the pixel spacing. Since some applications require pixel position errors less than the interpixel separation, to minimize these errors we must make the interface to the disk conform to this polar recording format. For example, to accurately record an image sensed by a television camera, the camera should be modified to scan along curved lines matching the shape of the tracks on the disk.

Although the optical disk is basically a binary storage medium, it can also encode grey level images. Area modulation can be used to code multiple reflectivity



Fig. 8. Grey scale image written on optical disk using area modulation.

levels for superpixels consisting of several bits. For example, turning on n out of N pixels in a superpixel can be used to represent the integer value n . Various superpixel coding techniques have been investigated in the past.⁸ We have implemented an area modulation scheme which uses a stochastic procedure to determine the position of on-pixel locations within each superpixel. In addition, this scheme improves the dynamic range of regions of low spatial frequency by stochastically selecting the value to be recorded in each element of an array of superpixels. Specifically, if a uniform region of p superpixels maps to a grey level between n and $n + 1$, randomly choosing between two levels for each superpixel in the region provides the entire p -superpixel region with an expected grey level equal to any one of $p - 1$ additional levels between n and $n + 1$. Figure 8 was generated using the area modulation recording scheme on the write-once disk and a schlieren imaging system as described above. The image shown consists of 512×480 superpixels each of dimension 6×4 pixels. The number of distinct grey levels recorded on the disk therefore was twenty-five. The grey levels present in the original image are clearly evident in the figure.

In addition to the recording and retrieval of images, the optical disk is an ideal medium for the storage of computer generated holograms (CGHs).^{9-12,25} The imaging technique described above may be thought of as simply the reconstruction of an image plane hologram. Any other computer generated hologram can just as easily be stored on the optical disk. We have investigated various techniques for the calculation and recording of CGHs on the optical disk.

The first method we investigated is based on using the computer to form the holograms of the individual points that make up an image. Thresholding a Fresnel

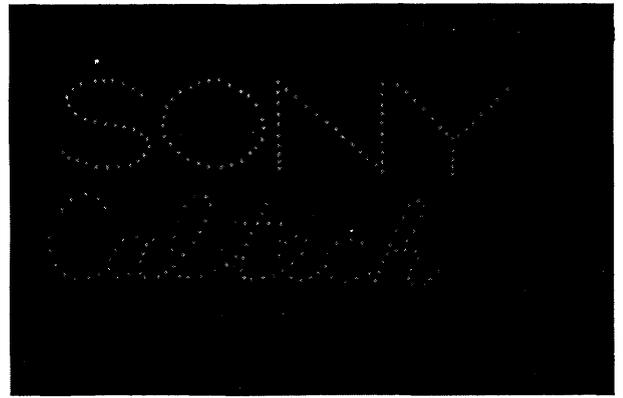


Fig. 9. Reconstruction of a binary Fresnel hologram stored on optical disk.

hologram of a single point (x_0, y_0, z_0) yields a Fresnel zone plate. It can be calculated by simply evaluating the real part of the Fresnel transform of a single point and thresholding the result, as shown below:

$$h_p(x, y) \propto \text{sgn} \left[\text{Re} \left\{ \iint \delta(x' - x_0, y' - y_0) \times \exp \left(j \frac{\pi}{\lambda z_0} [(x' - x)^2 + (y' - y)^2] dx' dy' \right) \right\} \right], \quad (20)$$

$$\propto \text{sgn} \left[\cos \left(\frac{\pi}{\lambda z_0} [(x - x_0)^2 + (y - y_0)^2] \right) \right]. \quad (21)$$

The reconstruction of the object point is achieved by illuminating the reflection hologram $h_p(x, y)$ with a plane wave. One component of the reflected field is a spherical wave converging to the point (x_0, y_0, z_0) . In a similar fashion, the Fresnel hologram of multiple points is calculated by summing many individual holograms and thresholding the result:

$$h(x, y) = \text{sgn} \left[\sum_m \cos \left(\frac{\pi}{\lambda z_m} [(x - x_m)^2 + (y - y_m)^2] \right) \right], \quad (22)$$

where (x_m, y_m, z_m) are the coordinates of the dots with which we construct the image that is stored holographically. The y_m terms are chosen in the range from 0 to $\lambda z_0 / \sqrt{\Delta_y^2 - \lambda^2}$, and the x_m terms in the range from 0 to $\lambda z_0 / \sqrt{\Delta_x^2 - \lambda^2}$. This guarantees that the reconstruction of the first-order in the y -direction does not overlap with any of the other orders. In effect, this is how we construct an off-axis hologram. If the number of dots that comprise the image is far less than the space-bandwidth product of the hologram, a reconstruction with high signal-to-noise ratio (SNR) is obtained.

Figure 9 shows the reconstruction of a binary Fresnel hologram we have recorded on a write-once disk. The reconstructed object contains over 100 points. The hologram consists of a 1000×1000 array of pixels recorded on the disk at a radius equal to 4 cm. By simply illuminating the hologram with a raw beam, we observe a self-focused reconstruction at a distance of 10 cm from the disk. The size of the reconstructed image is 3×5 cm which is much larger than the size of

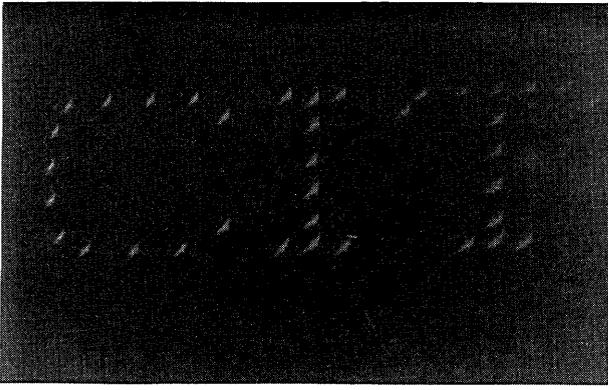


Fig. 10. Reconstruction of Fresnel hologram stored on magneto-optic disk.

the hologram itself (0.7×1 mm). This is a convenient feature which allows us to match the high density storage medium with a lower resolution array of detectors. Figure 10 shows the reconstruction of a Fresnel hologram recorded on the magneto-optic disk. In this case, the reconstructed hologram was observed through crossed polarizers to enhance the contrast. This polarization filtering technique, made possible by the polarization switching property of magneto-optic modulation, is necessary in this case because of the lower efficiency of the magneto-optic disk.

We can also record any conventional CGH (e.g., Lohmann or Lee).¹³⁻¹⁵ One CGH technique we have investigated consists of forming 1-D superpixels with position encoding of phase information. As can be seen in Fig. 11(a), each 1-D superpixel consists of four pixels along a specific track. By selecting the appropriate combinations of pixels to record within a group, each superpixel can be made to represent any of the nine phasors including the zero phasor shown in Fig. 11(b). We may also be able to represent a larger number of phasors without increasing the size of the superpixels by varying the size of the recorded pixels. In our system, this can be accomplished by either changing the energy used to write each pixel or by multiply exposing pixels.

We can also store Fourier holograms by simply recording the complex Fourier transform of the object field as described above. The Fourier transform hologram can be calculated in a number of ways; for exam-

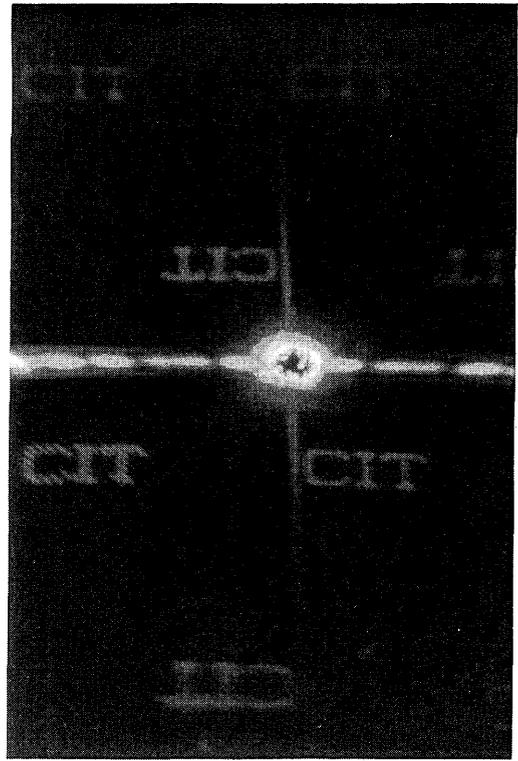


Fig. 12. Reconstruction of Fourier transform hologram.

ple, we could use the fast Fourier transform algorithm or we could calculate a Fresnel hologram using Eq. (22) with a large z_0 . This latter approach is equivalent to calculating the Fraunhofer diffraction pattern of the object and is the approach we have chosen. Shown in Fig. 12 is the reconstruction from a binary Fourier transform hologram. The data for this hologram were computed by thresholding the real part of the Fourier transform of the object. The hologram consists of 1000×1000 pixels written on the disk at a radius of 4 cm. Reconstruction was achieved using a Fourier transform lens.

One can also combine the Fourier and Fresnel hologram methods described above. For example, Fig. 13 shows the image reconstructed from a hologram generated using the Fourier transform in the along track direction and the Fresnel hologram in the across track direction. This transformation is given by the following equation:

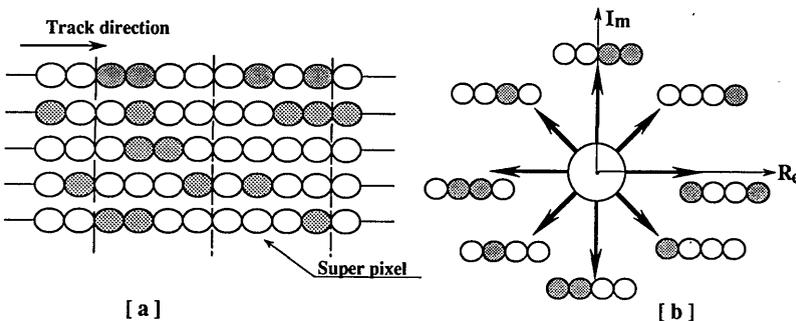


Fig. 11. Schematic of CGH technique used to record complex valued pixels: (a) one-dimensional superpixel used to code phase information; and (b) complex values recordable using (a).

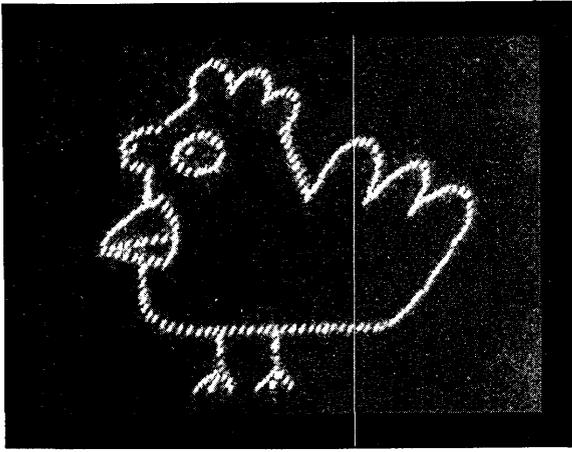


Fig. 13. Reconstruction of Fourier along track/Fresnel across track hologram.

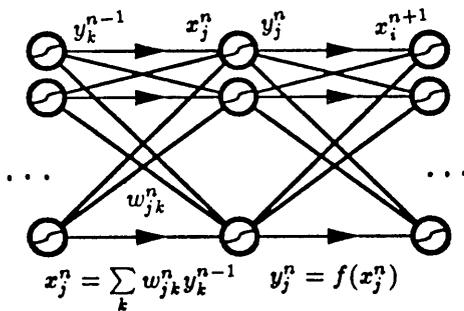


Fig. 14. Multilayer feedforward neural network.

$$h(x,y) = \iint b(x',y') \exp\left[j \frac{\pi}{\lambda z_0} (y - y')^2\right] \exp\left(-j \frac{2\pi}{\lambda z_0} x x'\right) dx' dy' \quad (23)$$

As before, we record a thresholded version of the real part of $h(x,y)$. Reconstruction is achieved using a cylindrical lens. Another storage scheme might involve calculating the 1-D Fourier transform of each line of the object and storing these 1-D signals along separate tracks. One benefit of using the Fourier transform in the along track direction and an imaging lens across track is that across track coherence is no longer required for parallel readout of 2-D data arrays since each track reconstructs a shift invariant line of data.¹⁶

We would also like to use the shift invariance property of the Fourier holograms in the along track direction to generate stationary reconstructions from a rotating disk. Unfortunately, this is not exactly true. As the disk rotates, the hologram experiences both translation and rotation. Rotation of a Fourier transform hologram results in an equal rotation of the reconstruction around the axis defined by the direction of the zero-order reflected beam. Thus, disk rotation results in apparent rotation and translation of the reconstructed image. We can neglect this motion if the reconstructed image moves less than a resolution element during the observation period. The size of a

resolution element in the reconstruction from an $N_x \times N_y$ -pixel Fourier transform hologram is given by

$$\begin{cases} \delta_x = \frac{\lambda F}{\Delta_x N_x} \\ \delta_y = \frac{\lambda F}{\Delta_y N_y} \end{cases} \quad (24)$$

in the x - and y -directions, respectively. F is the focal length of the Fourier transform lens used in the reconstruction.

The shift observed in the reconstruction plane as a function of disk rotation angle θ is

$$\begin{cases} e_x(\theta) = \frac{\lambda F}{2\Delta_y} [\cos(\theta) - 1] \\ e_y(\theta) = \frac{\lambda F}{2\Delta_x} \sin(\theta) \end{cases} \quad (25)$$

The hologram subtends an angle $\theta = N_x \delta_\theta$. In order that the reconstruction remain stationary during a rotation of the disk by θ , we find the following limits on the size of the hologram:

$$\begin{cases} N_x < \frac{\lambda F}{\Delta_x e_x(\theta)} \\ N_y < \frac{\lambda F}{\Delta_y e_y(\theta)} \end{cases} \quad (26)$$

Making a small θ approximation, we arrive at the following set of constraints for N_x and N_y :

$$\begin{cases} N_x < \frac{3\sqrt{4\Delta_y R}}{\delta_\theta} \\ N_x N_y < \frac{2R}{\Delta_y} \end{cases} \quad (27)$$

Thus, for the disk parameters in our experiment, the holograms must be smaller than 300×300 pixels, if we require the motion of the reconstruction to be negligible.

IV. Neural Networks

Neural network architectures are particularly well suited for the use of optical disks. This is due to the large memory that is typically required for the storage of the interconnecting weights in large networks. The optical disk not only supplies this storage capability, but in addition it provides the rapid access that is necessary for fast computation of the mapping of the network. Figure 14 shows a multilayer feedforward network, the most common neural network architecture. In this architecture, neurons are grouped in layers. The input to the j th neuron in the l th layer is x_j^l and its output is y_j^l . The weight of the interconnection between the input of this neuron and the output of the k th neuron in the previous layer is w_{jk}^l . Equation (28) governs the operation of the network:

$$\begin{cases} x_j^l = \sum_k w_{jk}^l y_k^{l-1} \\ y_j^l = f(x_j^l) \end{cases} \quad (28)$$

where f is typically a sigmoidal function.

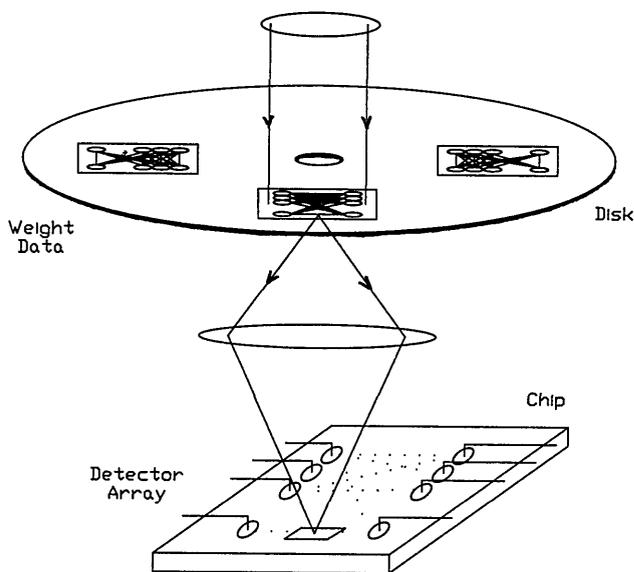


Fig. 15. Optical disk/VLSI hybrid neural network implementation.

An implementation of a feedforward network using optical disks combined with analog VLSI chips is shown in Fig. 15. The VLSI chip implements a single layer of the network. The weights that are used for propagating through one layer are optically loaded via the third dimension from the disk. An array of photodiodes is integrated into the VLSI chip for this purpose. The weights for the different layers of the network are stored adjacent to one another along the azimuthal direction of the disk. To achieve a multilayer network, we first download the weights for the first layer. The VLSI chip then evaluates the response of the first layer and stores the result. Meanwhile, the disk is spun so that the weights of the next layer are aligned with the chip, the new weights are downloaded and the response for the next layer is evaluated on the chip. This procedure is repeated until the response of the final layer is evaluated.

There are several possible implementations of the basic idea described in the previous paragraph. We describe a particular chip design which we are implementing experimentally. A schematic diagram of this chip is shown in Fig. 16. The neurons and synapses are arranged as a crossbar. The output of each neuron in the top row is a voltage source that raises the potential of its corresponding vertical wire to 5 V if the neuron is on or sets it to 0 V if the neuron is off. The amount of current that flows from each vertical wire into a horizontal wire is determined by the channel resistance of the field effect transistor FET at the corresponding intersection. The total current in each horizontal wire is the sum of the currents that were contributed from all the vertical wires. This summed current becomes the input to the neuron attached to each horizontal line. The neuron circuit accepts the input current, thresholds it, and generates an output voltage. The entire circuit is actually bidirectional. A horizontal wire that was previously used

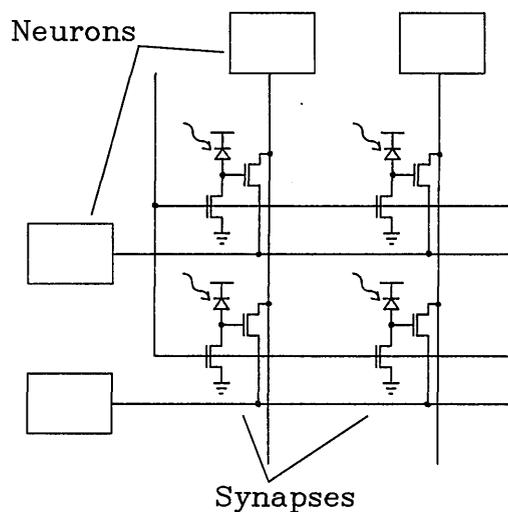


Fig. 16. Schematic of VLSI neural network crossbar implementation.

to sum the currents that flow from the vertical wires, can also be used to broadcast the state (voltage) of the neuron attached to it. Consequently, if the weights can be dynamically updated, a single chip can be used for the implementation of a multilayer network with data going back and forth between the two sets of neurons.

The strength of the connection between two neurons is increased by raising the gate voltage of the FET located at the intersection between the corresponding wires. The gate voltage is controlled by a circuit consisting of a series combination of a reverse-biased photodiode and a second, reset transistor. At the beginning of each cycle, the reset transistors are turned on, which sets the gates of the synapse transistors to 0 V. The reset transistors are then turned off for the rest of the cycle. The weights that are stored on the disk are imaged onto the chip. To facilitate this description, we will assume for the moment that the weights are binary (0,1) and we discuss later in more detail how to handle multivalued weights. When light from the disk strikes a photodiode on the chip, current flows through the photodiode charging the gate of the synapse transistor. This turns the transistor on, allowing current to flow between horizontal and vertical wires. We have not yet completed the testing of this chip, but we have characterized a circuit that consists of just two synapses that are however much more complex.

Figure 17 is a circuit diagram of one of the synapses that were fabricated. This synapse is a multiplying digital-to-analog converter (MDAC)¹⁷ contributing a current to the target neuron proportional to the product between the signal received from the input neuron and the weight, encoded as a binary quintuple, that is optically received by a set of five photodiodes. The weight and the neuron activity can be bipolar. This is achieved through the use of a pair of horizontal and vertical wires for each neuron. We denote the wires in the vertical and horizontal directions by V_m^+ , V_m^- and

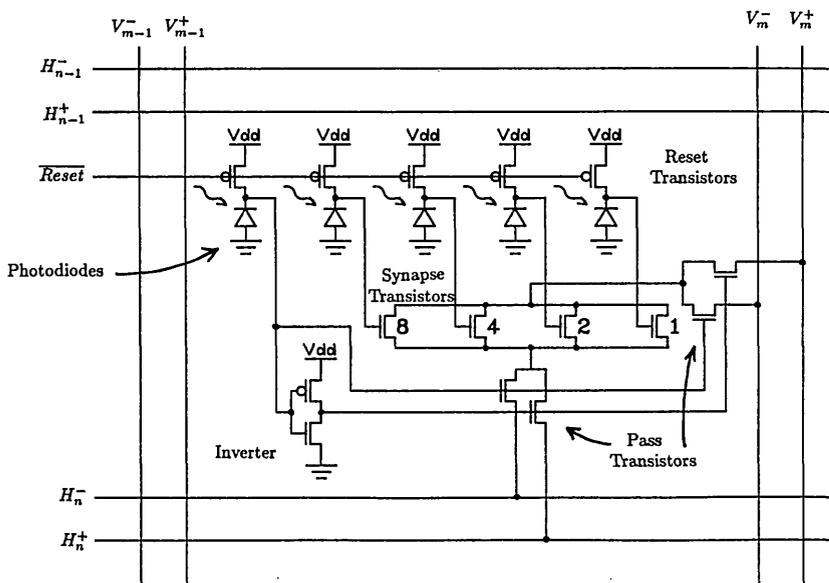


Fig. 17. Synapse circuit diagram.

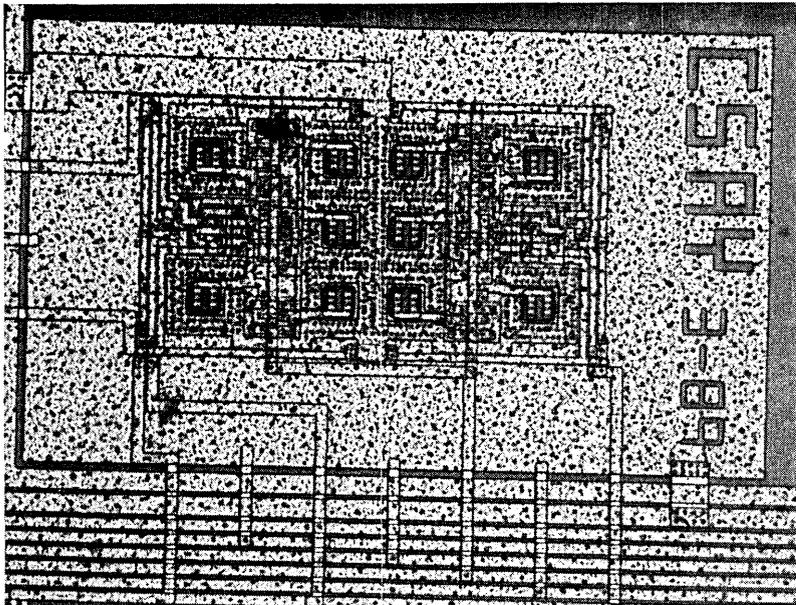


Fig. 18. Photograph of MDAC synapses.

H_n^+ , H_n^- , respectively. Four of the five photodiodes are dedicated to receiving the four bit weight value and the fifth, on the far left of Fig. 17, is for the sign bit. The five reset transistors in this design initially charge the gate voltages of the synapse transistors to V_{dd} . There are two pairs of pass transistors shown in Fig. 17. At any one time, one transistor of each pair is on and the other is off. In this manner, the setting of the pass transistors determines which two perpendicular wires are connected. The setting of the pass transistors is controlled by the sign bit with the help of the inverter circuit. The strength of the interconnection is determined by the setting of the synapse transistor, which in turn is determined by the optical signal on the photodiodes (the binary encoded weight). The channel resistance of an FET is proportional to its width-to-

length ratio (W/L). The analog-to-digital conversion is performed by scaling logarithmically W/L in the four synapse transistors. For example, suppose we are receiving a positive input signal along the vertical wires and the weight is negative. In this case, both vertical wires are set to 5 V by the output stage of the associated neuron. The sign bit through the pass transistors closes the circuit between the V_m^- and H_n^- wires. The amount of current that flows to H_n^- is proportional to the integer value corresponding to the four bit binary number detected by the photodiodes and also depends on the voltage setting ($V_{ref} \leq 5$ V) of H_n^- . If the sign bit flips, current will flow in the same direction but on H_n^+ . The target neuron subtracts the currents on H_n^+ and H_n^- to form its output and hence a bipolar weight is realized.

The photograph in Fig. 18 shows the pair of synapses which was fabricated through MOSIS. The circuitry for the synapses is protected by a layer of metal. Each of the ten squares in the figure containing π shaped features is a $26 \times 27\text{-}\mu\text{m}$ opening in this protective layer which allows light to strike the $20 \times 21\text{-}\mu\text{m}$ photodiode located underneath. The plot of Fig. 19 shows the current on H_n^+ vs V_{ref} for voltages of 5 V (neuron on) and 0 V (neuron off) on V_m^+ . When V_{ref} is set to 0 V (5 V), the resulting current swings between 0 μA ($-140 \mu\text{A}$) and 140 μA (0 μA). In other words, when V_{ref} is set to either 0 V or 5 V, this is a two-quadrant multiplier. If, however, V_{ref} is set to any intermediate value, the output current swing is bipolar. In particular, for the circuit we fabricated, when $V_{\text{ref}} = 1.17$ V, the positive and negative deviations of the current are symmetric and we have a balanced four-quadrant multiplier.

The speed of the neural network chip is limited principally by the time required to discharge the gates of the synapse transistors through the photodiodes. The discharge current depends on the amount of light striking it and the voltage across the photodiode. Figure 20 shows the current flowing through the synapse as a function of time as all the photodiodes are simultaneously struck with light for three different intensity levels; V_{ref} was set to 1.17 V for this measurement. As expected, the response time is inversely proportional to the light intensity and corresponds to a switching energy of 1.55 pJ. This chip was not optimized for speed (phototransistors could be used in place of the photodiodes); however, we can use the measurement of Fig. 20 to obtain an estimate for the speed with which we can operate such a network. Let us assume that we have available optical power equal to 1 W and a chip consisting of 100×100 synapses. The diffraction efficiency of the disk was estimated to be approximately 0.1% which yields 1 mW total power incident on the chip, or 100 nW per synapse. Dividing the switching energy by the available power per synapse, we obtain 15.5 μs response time. This corresponds to 3 Gbits/s transfer rate between memory and chip. Even though this is a remarkable rate, there is a lot of room for improvement in the speed in this design, through a disk with better light efficiency, the use of phototransistors, and a reduction in the number of bits used in the MDACs.

In the neural network implementation described above, the disk is used only as a parallel read-out storage device and the electronic chip is used to perform all the calculations. It is also possible to use the disks as computer-generated holograms or transparencies in many of the optical neural network architectures that have been previously proposed.¹⁸ In such optical neural networks, the analog multiplication needed to implement the weights is performed by propagating an incident light field through a transparency (the disk in this case) and summing multiple such products onto a single detector location. To accomplish this, the output of each neuron must be an optical signal. This can be done through the use of spatial

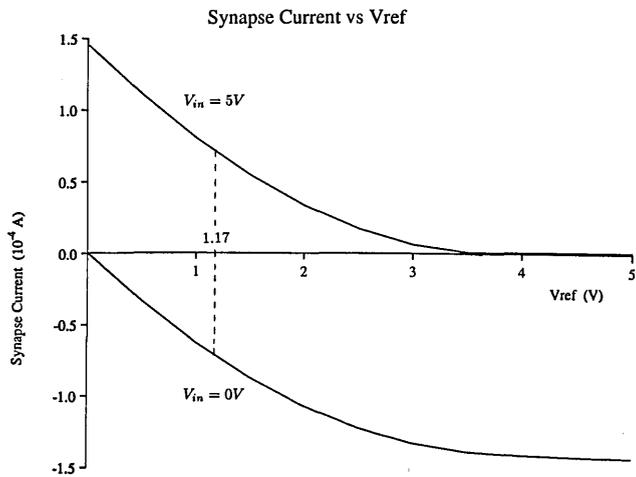


Fig. 19. Synapse current dependence on V_{ref} .

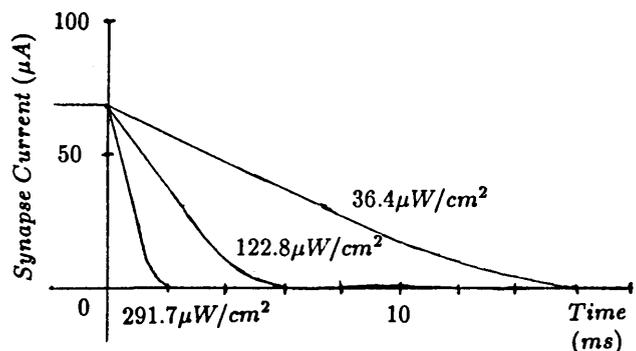


Fig. 20. Synaptic time response dependence on light intensity.

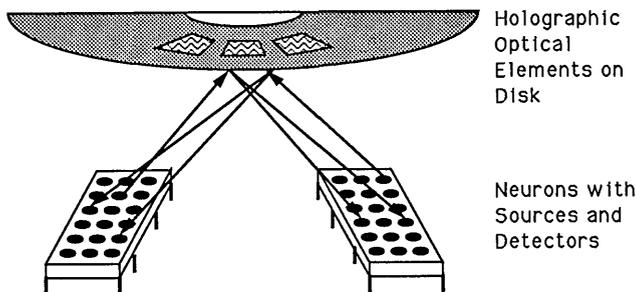


Fig. 21. Optical disk implementation of synaptic connections.

light modulators or optoelectronics. The optoelectronic approach for building the neurons combined with implementation of the synapses on the disk is schematically shown in Fig. 21. The neurons can be fabricated in GaAs¹⁹ on which detectors, sources, and electronic circuits can be monolithically integrated. We can greatly increase the density of neurons on the GaAs chip compared to the previous approach, since we no longer require circuitry on the chip to detect the weights and multiply them with the neuron outputs. Second, since optics provides us with greater flexibility in performing the interconnections between the neu-

rons, we can, at least in principle, construct not only larger but also more general neural network architectures beyond multilayer feedforward networks.

V. Correlators

The correlation function $c(\tilde{x}, \tilde{y})$ is defined as

$$c(\tilde{x}, \tilde{y}) = \mathcal{F}^{-1}\{F(w_x, w_y)G^*(w_x, w_y)\} \quad (29)$$

$$= \iint f(x, y)g(x - \tilde{x}, y - \tilde{y})dx dy, \quad (30)$$

where $f(x, y)$ and $g(x, y)$ are two real images, $F(w_x, w_y)$ and $G(w_x, w_y)$ are their respective 2-D Fourier transforms, and $\mathcal{F}^{-1}\{\}$ is the 2-D inverse Fourier transform operator. It is well known that $c(\tilde{x}, \tilde{y})$ is sharply peaked at the point (x_0, y_0) when $f(x, y) = g(x - x_0, y - y_0)$. This property is what makes the correlation function useful for pattern recognition because, regardless of the position of the input image $g(x, y)$, $c(\tilde{x}, \tilde{y})$ will have a peak if $f(x, y)$ and $g(x, y)$ are matched. Since, in general, there are many versions of an image g that we would like to recognize, a reliable image recognition system should provide invariance to multiple object attributes. Often the best way to achieve this invariance is to use a large number of reference patterns f against which to compare g in order to obtain reliable recognition.

Optical image correlators based on Fourier transform (FT) holograms were proposed by VanderLugt in 1964.²⁰ For optical correlation to be a realistic approach to image recognition, we require a memory device sufficient to store a large reference image library, an SLM which interfaces with this memory in real time, and a scanning or addressing mechanism which allows interrogation of the entire reference library in a reasonable amount of time. The optical disk provides these three characteristics in one device. In this section, we describe several optical disk based image correlation architectures and present experimental results taken from selected systems.^{21,22} We will examine critical parameters associated with each architecture and evaluate each system in terms of power and speed.

A. VanderLugt CGH Correlator

The first disk based image correlator to be described is the simple VanderLugt correlator shown in Fig. 22. As can be seen from the figure, a Fourier transform computer-generated hologram recorded on the optical disk is used as a Fourier plane filter for the input image. The product of the transforms of the input and reference images is formed at the disk and an inverse transform yields the desired 2-D correlation in the output plane. As the disk rotates, a new correlation pattern is generated every time a different CGH aligns with the input image FT. Therefore, whenever there is a match between the input FT and the CGH, a peak occurs in the output plane of the system. The location of this peak, which corresponds to the location of the object of interest in the input plane, may be anywhere in the correlation plane. Therefore, a 2-D detector array is required to acquire the correlation data. Fur-

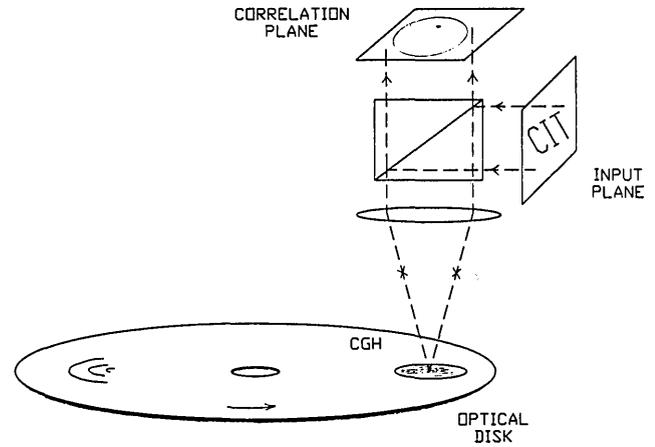


Fig. 22. Optical disk based VanderLugt correlator.

thermore, since the correct correlation only occurs during the brief periods of alignment between the input FT and the CGH, the detector array need only be queried at these times. A pulsed laser or an electronically gated detector array could be used to achieve the appropriate sampling. The proper operation of this system depends on the optical quality of various disk coating materials which are capable of introducing random phase distortions across the system filter plane. As we discussed in Sec. II, however, the optical quality of several commercially available disks is sufficient to make them suitable for these applications.

As with other FT based architectures, one advantage of this correlator is its potentially high speed. The correlation rate in this system is limited by disk rotation rate and detectability of the correlation peak. Taking a typical disk rotation rate of 40 Hz, we calculate a correlation rate $R_c = 400,000$ correlations/s for 100×100 pixel images. This correlation rate implies that to detect a correlation peak, each element of the 2-D detector array in the output plane must have a bandwidth of ≈ 400 kHz. Further, we can calculate the peak detectability n_p given by the number of photons detected at the correlation peak, by first calculating the peak dwell time τ , multiplying this by the expected power in the correlation peak P_c , and dividing by the photon energy. That is:

$$n_p = \tau P_c / h\nu, \quad (31)$$

where ν is the frequency of the optical field. Peak dwell time is given simply by $\tau = 1/NR_c$ where N is the number of pixels in the reference image in the along track direction. In the case of random, bipolar images, the expected fraction of diffracted power that will be measured at the peak is $1/2$. By random we mean that each pixel of the image is equally likely to take on either of the two possible values ± 1 . Given the disk diffraction efficiency, η , and the source power P_s , the correlation peak power is

$$P_c = \eta P_s / 2. \quad (32)$$

Substituting $P_s = 10$ mW and $\eta = 0.1\%$, we find an expected peak power $P_c = 5 \mu\text{W}$. This rate leads to an

easily detectable 10^5 photons in the correlation peak. The two most significant drawbacks of this system are alignment criticality and computational overhead. For each reference filter, a 2-D FT CGH must be computed and written on the disk. For a large reference library the time required for this procedure can be long. More importantly, the alignment of the input FT and the CGH is critical to within the resolution of the CGH ($\approx 1 \mu\text{m}$). As the disk rotates, nonuniformities resulting from wobble and disk center offset, lead to nonuniformities in reference image location with respect to the optical system. For example, the Sony WORM disks we use allow up to 1° of wobble and up to $50 \mu\text{m}$ offset between the rotational center of the disk and the actual track center. These nonuniformities must be compensated for the output correlation to be accurate.

B. Photorefractive Correlator

Since the wobble and offset problems introduce slowly varying nonuniformities (<50 wobble cycles/rotation), the problem of alignment sensitivity can be effectively dealt with using real time compensation with feedback of the sort used in commercial disk drives; however, the computational overhead associated with generating the desired reference library in the above system remains a problem. The system of Figure 23 eliminates this processing time by allowing the reference images themselves to be recorded on the disk instead of FT CGHs. In this system, a photorefractive crystal or any other real time, temporary holographic storage medium is used to record a hologram of the input FT. During the recording phase, the disk illumination is blocked and the input transparency is illuminated from the right. A hologram is formed between the input FT and the reference beam as shown. This hologram will then be read out using the reference library. On readout, the input is blocked and the disk is illuminated. The product of the input and reference FTs is formed in the crystal and inverse transformed to yield the correlation output.

If the photorefractive crystal is replaced by a thin medium such as a holographic plate, then the output pattern is exactly the desired 2-D correlation; however, it has been shown that when a thick hologram is used in the filter plane of such a system, the resulting output is a 1-D slice of the 2-D correlation pattern.²³ This can be understood by considering the recording arrangement shown in Fig. 23. On recording, each plane wave corresponding to one of the points in the input image forms a grating with the reference beam. The resulting hologram exhibits Bragg selectivity in the horizontal direction. 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 array of holograms and generates a vertical array of spots in the correlation plane at the horizontal location corresponding to the reference beam FT. The coherent sum of all such reconstructions comprises the output of the correlator. This output pattern is the desired 2-D

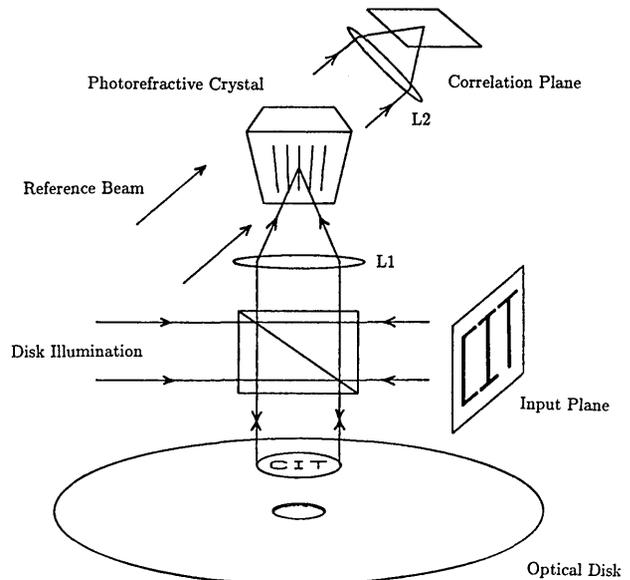


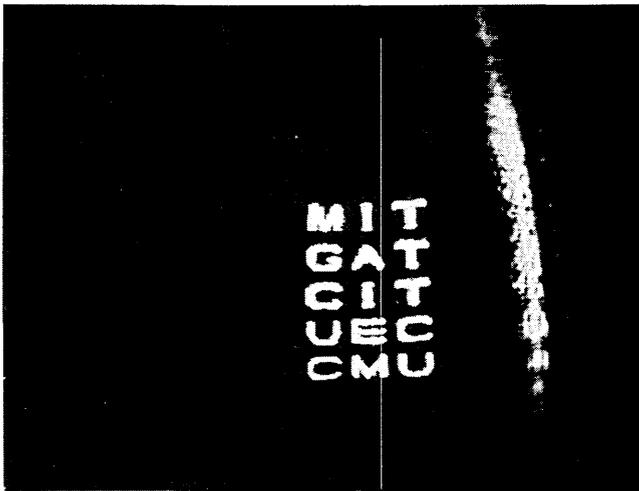
Fig. 23. Photorefractive/optical disk based correlator.

correlation multiplied, in the horizontal direction, by a sinc function whose width is inversely proportional to the hologram thickness. In the system of Fig. 23, this property does not cause problems since all 1-D slices are obtained sequentially via disk rotation. Further, instead of requiring a full 2-D detector array at the output, a 1-D array is sufficient to sequentially detect each slice of correlation output. Despite the advantages gained in terms of computational overhead and detector simplicity, alignment compensation remains a critical issue with this system. The expected correlation rate obtainable using this system is again limited primarily by disk speed and peak detectability. A rate of 400,000 correlations/s is still easily achievable, yielding a detected signal at the correlation peak of more than 10^5 photons using again a 10 mW source.

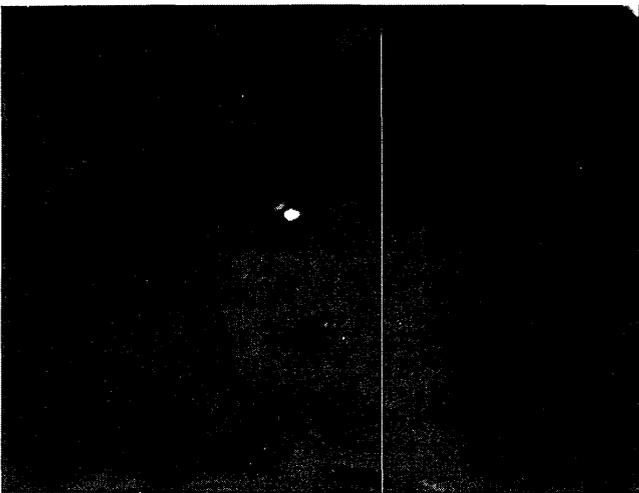
This system has been experimentally demonstrated using a thin hologram in place of the photorefractive crystal. For this experiment, we chose to use a Sony WORM disk as an SLM at the input as well as the reference. The results are shown in Fig. 24. Figure 24(a) shows the image recorded on two Sony disks; one disk was used to record the filter plane hologram on the plate and the another was used to read out the hologram. The correlation plane output is shown in Fig. 24(b). The characteristic autocorrelation peak appears in the output.

C. Rotating Mirror Correlator

The previous two Fourier transform based correlators implement Eq. (29) to generate the correlation function. While FT based systems are typically superior in terms of speed, alignment and coherence requirements are relative disadvantages. We discuss next two systems which perform 2-D correlations based on Eq. (30). In these systems the correlation function is generated by calculating an inner product for every relative shift between input and reference



(a)



(b)

Fig. 24. Disk based Fourier plane correlator results using a plate: (a) input image recorded on disk 1 and used to record the hologram; (b) correlation pattern obtained using disk 2 as reference to read out the hologram.

images. Since these shifts will be generated sequentially, the correlation will appear as a 1-D signal representing a raster version of the desired 2-D correlation pattern. As we will see, these systems sacrifice correlation rate for operational simplicity without alignment criticality, while at the same time relieving source coherence requirements.

A simple image plane correlator is shown in Fig. 25. An image of the input scene is formed at the disk on which a library of reference images resides. The total transmitted or reflected light is collected by a detector at the output. The rotation of a polygon mirror causes the input image to scan the disk radially while the disk rotation itself provides scanning in the orthogonal direction. The detected light therefore represents the instantaneous inner product between the input and a shifted version of one of the references. All relative shifts between input and reference images are generat-

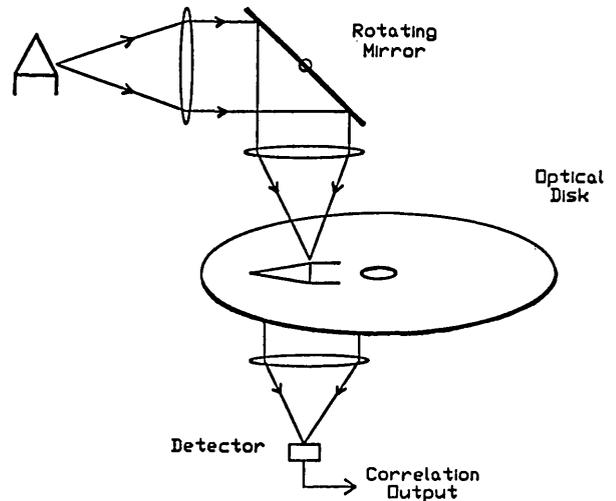
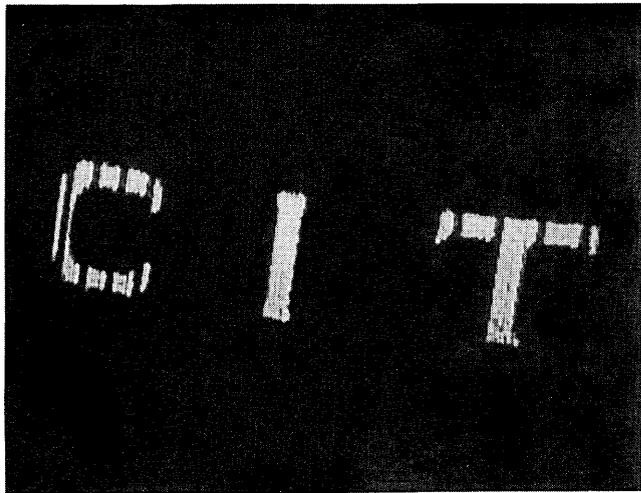


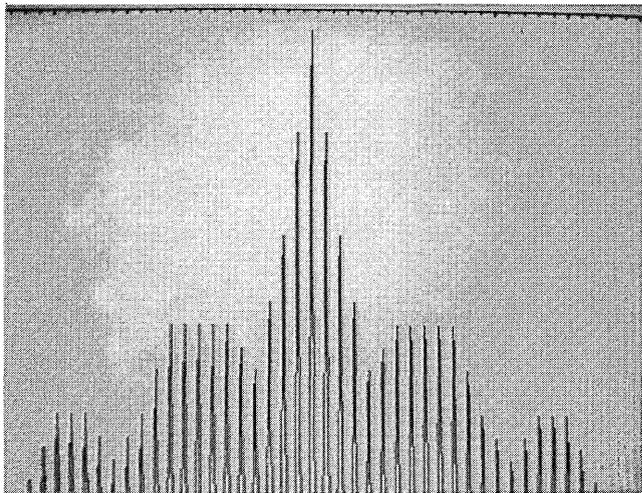
Fig. 25. Rotating mirror correlator.

ed using these two scanning mechanisms and as a result, the light collected at the output is exactly a 1-D raster signal of the desired 2-D correlation. An example of the output obtained from the rotating mirror correlator is shown in Fig. 26 along with a computer simulation of the desired 2-D autocorrelation function displayed as a 1-D raster. Figure 26(a) is the reference image written on a Sony write once disk. This image has up to 6912 pixels along track and comprises 1024 tracks. The input to the system was provided by a transparency of the acronym CIT illuminated by a He-Ne laser. Figure 26(c) shows the correlation signal generated by the optical system for this input. The asymmetry in the optical autocorrelation is due to a slight mismatch between the input and reference images.

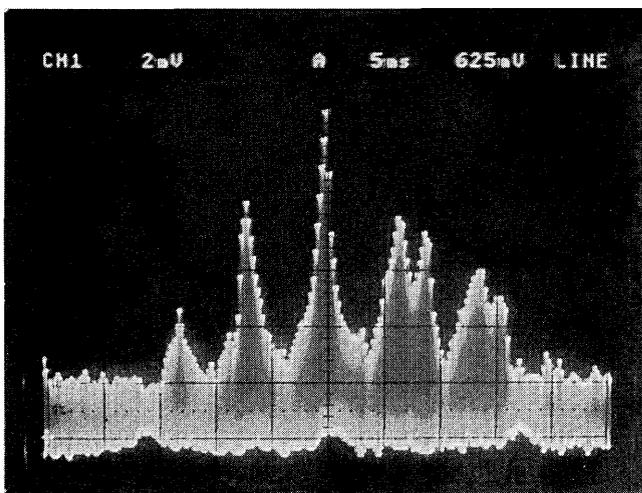
This system is capable of operating with incoherent illumination. One critical limitation, however, is its speed. The speed in this architecture is dictated primarily by the rates of the relative scanning mechanisms. Specifically, to generate an accurate correlation signal, the radial scan time must be less than the time it takes the references to rotate by 1 pixel. The correlation rate is thus limited by factors such as reference image, pixel size, radial scan rate, disk rotation speed, and ultimately by illumination level and disk efficiency. For our experiment, the speed of the rotating mirror is the limiting factor, and it results in a correlation rate of 400 correlations/s for 100×100 pixel images. This system, although significantly slower than FT based systems, provides a simple solution to the FT correlator alignment problems by operating in the image plane. As with any incoherent correlator, the present system is a unipolar architecture and some bias removal mechanism is necessary to retrieve a bipolar or high SNR correlation signal. These bias removal techniques have been discussed at length in the literature.²⁴



(a)



(b)



(c)

Fig. 26. Rotating mirror correlator results: (a) reference image recorded on Sony disk; (b) computer-generated autocorrelation signal; and (c) optical system output.

D. Acousto-optic Correlator

The most obvious way to improve the speed of the above system is to increase the speed of the radial scanning mechanism. Since the speed of commercially available polygon mirror based scanners is limited to about 40 kHz, we have considered instead the system of Fig. 27 which utilizes an acousto-optic (AO) device as the radial scanner. The AO scanner can achieve scan rates up to 10 MHz. In this architecture, a chirp signal propagating in the AO device generates a moving cylindrical lens with power in the horizontal dimension. This moving cylindrical lens becomes part of the system that images the input onto the disk. Consequently, as the AO lens moves horizontally, the image formed on the disk is scanned radially. The orthogonal scanning is achieved by disk rotation as before and the light collected by the detector once again represents the desired correlation signal.

The correlation rate in this system is still constrained by the radial scan speed; however, since this scanning is generated by virtue of the propagation of a RF chirp in the Bragg cell, the resulting correlation rate is much higher than before. The RF chirp parameters are chosen so as to utilize as much AO space-bandwidth product as the input image requires, while minimizing scan time. Specifically,

$$t_C = t_{AO}(SBP_{IN}/SBP_{AO}), \quad (33)$$

where t_C is the required RF chirp duration, t_{AO} is the AO aperture, and SBP_{IN} and SBP_{AO} are the input and AO space-bandwidth products, respectively. We have built this system using a TeO_2 AO cell with a $70 \mu\text{s}$ aperture and a RF chirp centered at 40 MHz with a chirp rate of $\approx 4 \text{ MHz}/\mu\text{s}$. Using the above equation with $SBP_{IN} = 100$ and $SBP_{AO} = 1000$ the required chirp duration was calculated to be $7 \mu\text{s}$. A SAW device was used to generate the desired chirp signal. The resulting radial scan rate of $1/7 \mu\text{s} \approx 140 \text{ kHz}$, yields a correlation rate of 1400 correlations/s. Again, this correlator is incoherent, but quasimonochromatic light would be required because of the wavelength sensitivity of the AO lens. The impulse response of the AO lens scanner is shown in Fig. 28. The image of an input transparency was formed on a CCD using one cylindrical lens with power in the horizontal dimension and the AO lens for vertical imaging. The input illumination was pulsed so that the AO lens might be frozen in various vertical positions. The delay between the onset of the RF chirp and the laser diode pulse determines the position of the image on the CCD. As can be seen from the figure, the AO lens imaging characteristics are quite good. The output of the AO lens correlator is shown in Fig. 29. The input to this system was once again the transparency of the acronym CIT and the reference was a duplicate CIT written on an Optotech WORM disk using a simple recording system which we built. The correlator output is, therefore, the 2-D autocorrelation of the input image. The reference image shown in Fig. 29(a) is relatively large so that a radial scanning distance of 2 cm was

Input Plane

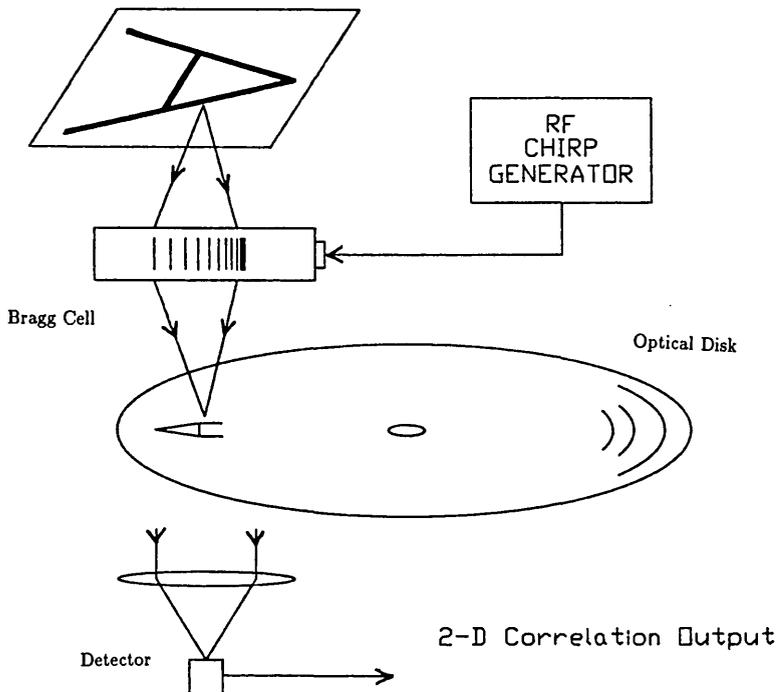


Fig. 27. Moving AO lens correlator.

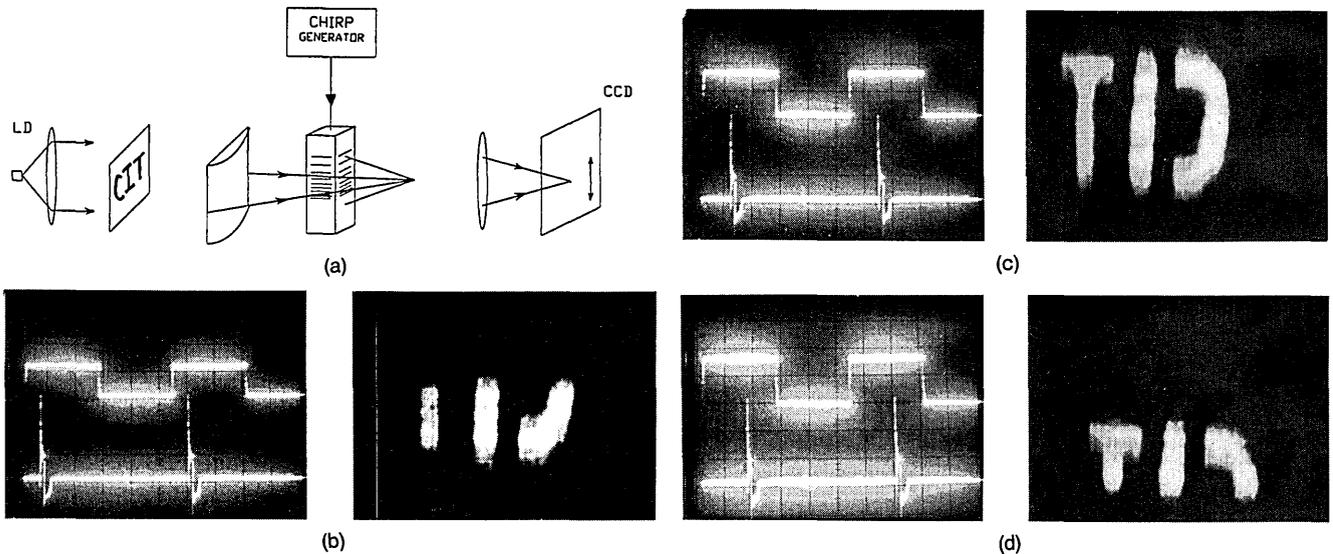


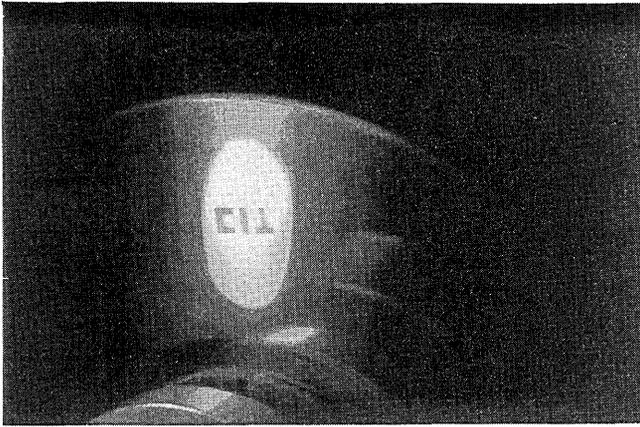
Fig. 28. Impulse response of the AO lens scanner: (a) optical system used to measure impulse response; (b)–(d) image formed on CCD for various delay times Δ , where Δ is the time between the leftmost edge of the chirp gate (upper trace) and the laser diode trigger (lower trace), (b) $\Delta = 2.5 \mu\text{s}$, (c) $\Delta = 5.0 \mu\text{s}$, (d) $\Delta = 8.0 \mu\text{s}$.

required to generate an accurate correlation signal. As can be seen from Figs. 29(b) and 29(c), the optical system output agrees well with the predicted autocorrelation signal of Figure 26(b).

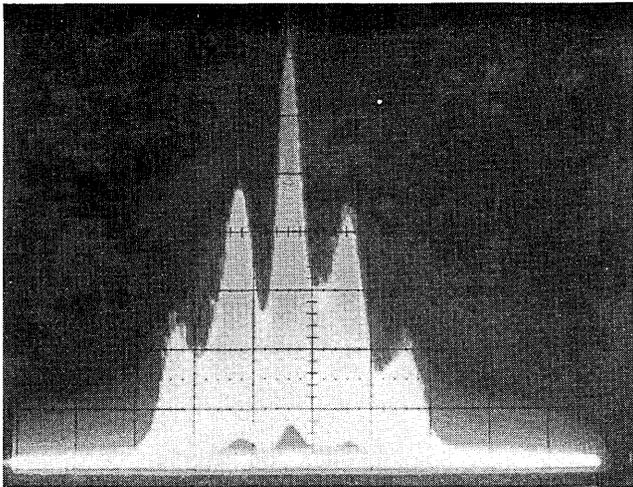
VI. Conclusions

This paper describes the use of optical memory disks in optical computing and optical information process-

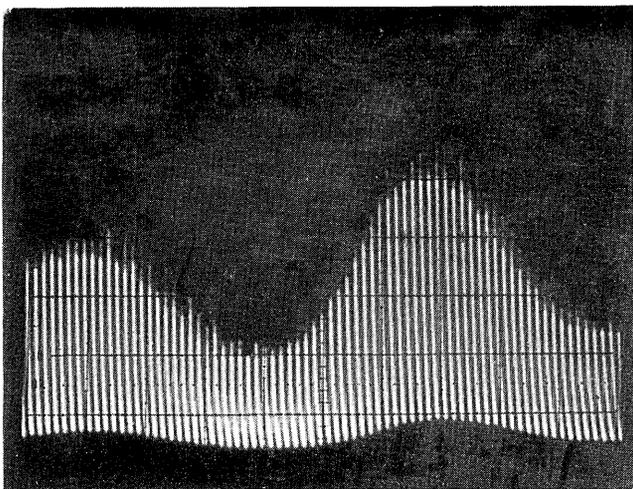
ing systems. The large SBP ($\approx 10^{10}$), simple computer addressability, natural scanning mechanism, and parallel accessibility are all features making the optical disk a candidate for use as both memory and SLM in these systems. The sampled format recording scheme results in across track coherence, facilitating the storage of 2-D data on the disk. This format also results in the absence of grooves on the disk, which eliminates



(a)



(b)



(c)

Fig. 29. Output of AO lens correlator: (a) reference on disk; (b) optical system output; and (c) magnified version of (b) to resolve individual radial scan peaks.

sampling of the bias reflectivity thereby increasing image plane contrast. The third attractive characteristic of the disks we used is the optical flatness of the glass coating material. This allows these disks to be used in coherent processing such as holographic reconstruction and complex spatial filtering.

Parallel optical access to images and holograms stored on disks provides the possibility of implementing specialized parallel computing schemes such as database machines, image correlators, and optical disk/VLSI hybrid neural networks. We have demonstrated several of these systems and have shown the potential advantage of such systems over their electronic counterparts. Owing to the maturity of optical disk technology, the architectures described here are feasible using existing disk systems and readily available supporting devices.

This work is supported in part by a grant from the Army Research Office. The authors would like to thank Sony for the disk system used in this work and for the generous support of Seiji Kobayashi during his stay at California Institute of Technology. Special thanks to Adolf Lohmann for many helpful discussions regarding optical disk based computer generated holography. Thanks to Charlie Stirk for his assistance in the design and testing of our VLSI neural net chip. Alan Yamamura is supported by a fellowship from the Fanny and John Hertz Foundation.

Appendix: Diffraction Efficiency for Schlieren Imaging

We calculate, as an example, the diffraction efficiency of the optical disk in a schlieren imaging system. We start by modeling the disk surface according to the following equation:

$$\begin{aligned}
 i(x,y) = & r_0 + (r_1 - r_0) \left[b(x,y) \sum_{n,m} \delta(x - n\Delta_x, y - m\Delta_y) \right] \otimes \text{circ}\left(\frac{r}{\Delta_r}\right) \\
 & + (r_2 - 2r_1 + r_0) \sum_{n,m} \left\{ [b(x,y)\delta(x - n\Delta_x, y - m\Delta_y)] \otimes \text{circ}\left(\frac{r}{\Delta_r}\right) \right\} \\
 & \times \left\{ [b(x,y)\delta(x - (n+1)\Delta_x, y - m\Delta_y)] \otimes \text{circ}\left(\frac{r}{\Delta_r}\right) \right\}, \quad (A1)
 \end{aligned}$$

where $r^2 = x^2 + y^2$, $b(x,y)$ is the desired binary image, Δ_x the along and Δ_y the across track spacing, Δ_r the radius of the written spots, and r_i the complex amplitude reflectivity of areas written i times. Because spots written by the Sony system have a constant angular separation along track, Δ_x is actually a function of radial position on the disk and varies between Δ_r and $2\Delta_r$. The third term in Eq. (A1) represents the overlap of adjacent spots in the along track direction when $\Delta_x < 2\Delta_r$ (Fig. 30). We otherwise ignore track curvature for now and assume that the pixels lie on a Cartesian grid. The effects of track curvature are analyzed in Secs. II and III.

Because the Sony write once material nearly saturates after a single exposure to the write beam such that $r_2 \approx r_1$, we drop the third term of Eq. (A1) and

account instead for the overlap by modifying the spot shape in the second term by assigning half of the overlap region to each neighboring spot, as in Fig. 30. In this case, the mathematical expression for the spot becomes [circ \times rect] and Eq. (A1) reduces to the following:

$$i(x,y) = r_0 + (r_1 - r_0) \left[b(x,y) \sum_{n,m} \delta(x - n\Delta_x, y - m\Delta_y) \right] \otimes \left[\text{circ}\left(\frac{r}{\Delta_r}\right) \text{rect}\left(\frac{x}{\Delta_x}\right) \right]. \quad (\text{A2})$$

Strictly speaking, the above equation models the reflectivity incorrectly wherever a written spot is adjacent along track to an unwritten one. Assuming, however, that along track spatial frequencies in the image are low compared to the sampling frequency Δ_x^{-1} , we expect separate clusters of written and unwritten spots, in which case there would be little energy in a term accounting for written spots next to unwritten ones.

Using Eq. (A2) to model the reflectivity of the disk surface, we find the following Fraunhofer diffraction pattern:

$$I(u,v) = r_0 \delta(u,v) + (r_1 - r_0) \left[B(u,v) \otimes \frac{1}{\Delta_x \Delta_y} \sum_{n,m} \delta\left(u - \frac{n}{\Delta_x}, v - \frac{m}{\Delta_y}\right) \right] \times \left[\Delta_r \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right]_{\rho^2 = u^2 + v^2}. \quad (\text{A3})$$

As expected, the sampling in the image plane corresponds to convolving the image spectrum with an array of impulses resulting in an array of image spectra or diffraction of the image into multiple orders. Schlieren imaging achieves high contrast because the pixels sample only the image and not the background, represented by r_0 , sending all the energy in the bias to the zero-order.

The fraction of incident light that goes into the n,m th diffraction order is given by integrating the magnitude squared of the appropriate term from Eq. (A3) as follows:

$$H_{nm} = \left| \frac{\Delta_r(r_1 - r_0)}{\Delta_x \Delta_y} \right|^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| B\left(u - \frac{n}{\Delta_x}, v - \frac{m}{\Delta_y}\right) \right|^2 \times \left[\frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right]^2 dudv. \quad (\text{A4})$$

The shape of the written spots determines the characteristic $[J_1 \otimes (\text{sinc} \times \delta)]$ envelope which modulates the entire diffraction pattern in Eq. (A3). Assuming that the pixels sufficiently oversample the input image, the envelope is nearly constant over the image spectrum allowing us to simplify Eq. (A4):

$$H_{nm} \approx \left| \frac{(r_1 - r_0)}{\Delta_x \Delta_y} \right|^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \left(u - \frac{n}{\Delta_x}, v - \frac{m}{\Delta_y}\right) \right|^2 dudv \times \left| \Delta_r \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right|^2_{u=n/\Delta_x, v=m/\Delta_y}. \quad (\text{A5})$$

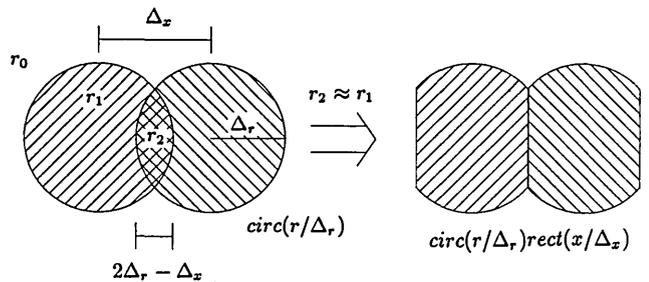


Fig. 30. Model of spot shapes.

The remaining integral corresponds to the energy in the spectrum of the image which is equivalent to the energy in the image itself. Thus, the fraction of incident light energy that goes into the n,m th diffraction order is given as follows:

$$H_{nm} \approx \left| \frac{r_1 - r_0}{\Delta_x \Delta_y} \right|^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |b(x,y)|^2 dx dy \times \left| \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right|^2_{u=n/\Delta_x, v=m/\Delta_y}. \quad (\text{A6})$$

Defining useful light as the total energy going into image spectra of all orders, we estimate the light captured by imaging a single order as a fraction of useful light:

$$\frac{H_{nm}}{\sum_{n',m'=-\infty}^{\infty} H_{n'm'}} = \frac{\left| \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right|^2_{u=n/\Delta_x, v=m/\Delta_y}}{\sum_{n',m'=-\infty}^{\infty} \left| \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right|^2_{u=n'/\Delta_x, v=m'/\Delta_y}}. \quad (\text{A7})$$

Since the denominator is the sum of the squares of the Fourier series coefficients of an image with every spot written with unity amplitude, it is equivalent to the fraction of the disk covered by spots if all were written. This fraction varies between 0.877 at the innermost radius and 0.785 at the outermost. We can now simplify Eq. (A7) as follows:

$$\frac{H_{nm}}{\sum_{n',m'=-\infty}^{\infty} H_{n'm'}} = \frac{\left| \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right|^2_{u=n/\Delta_x, v=m/\Delta_y}}{\frac{1}{\Delta_x \Delta_y} \int_{-\Delta_x/2}^{\Delta_x/2} \int_{-\Delta_y/2}^{\Delta_y/2} \left| \text{circ}\left(\frac{r}{\Delta_r}\right) \text{rect}\left(\frac{x}{\Delta_x}\right) \right|^2 dx dy} = \frac{\left| \frac{J_1(2\pi\Delta_r\rho)}{\rho} \otimes \Delta_x \text{sinc}(\Delta_x u) \delta(v) \right|^2_{u=n/\Delta_x, v=m/\Delta_y}}{\frac{1}{\Delta_x \Delta_y} \left[\frac{2\Delta_r^2}{\Delta_x \Delta_y} \sin^{-1}\left(\frac{\Delta_x}{2\Delta_r}\right) + \frac{\Delta_r}{\Delta_y} \sqrt{1 - \frac{\Delta_x^2}{4\Delta_r^2}} \right]}. \quad (\text{A8})$$

Thus, Eq. (A6) shows the amount of light as a fraction of incident light that goes into the image spectrum

in each order. Equation (A9) shows the amount of image light in each order as a fraction of image light in all orders. These equations can be used to estimate the diffraction efficiency of the Sony write-once disk in a schlieren imaging system. The disk diffraction efficiency for other applications and/or disks can be estimated in a similar fashion given a model for the reflection or transmission pattern of the disk and parameters for the appropriate disk characteristics.

References

1. R. Bartolini, H. Weakliem, and B. Williams, "Review and Analysis of Optical Recording Media," *Opt. Eng.* **15**, 99-108 (1976).
2. Y. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," *Sci. Am.* **255**, 88-95 (1987).
3. D. Psaltis, A. A. Yamamura, M. A. Neifeld, and S. Kobayashi, "Parallel Readout of Optical Disks," in *Technical Digest, Topical Meeting on Optical Computing* (Optical Society of America, Washington, DC, 1989), pp. 58-61.
4. L. Giles and B. K. Jenkins, "Models of Parallel Computation and Optical Computing," in *Technical Digest, OSA Annual Meeting* (Optical Society of America, Washington, DC, 1986), paper ML1.
5. Y. Nakane *et al.*, "Principle of Laser Recording Mechanism by Forming an Alloy in the Multilayer of Thin Metallic Films," *Proc. Soc. Photo-Opt. Instrum. Eng.* **529**, 76-81 (1985).
6. D. Psaltis, E. G. Paek, and S. S. Venkatesh, "Optical Image Correlation with a Binary Spatial Light Modulator," *Opt. Eng.* **23**, 698-704 (1984).
7. J. H. Rilum and A. R. Tanguay, Jr., "Utilization of Optical Memory Disks for Optical Information Processing," in *Technical Digest, OSA Annual Meeting* (Optical Society of America, Washington, DC, 1988), paper M15.
8. J. F. Jarvis, C. N. Judice, and W. H. Ninke, "A Survey of Techniques for the Display of Continuous Tone Pictures on Bilevel Displays," *Comput. Graphics Image Process.* **5**, 13-40 (1976).
9. Y. Tsunoda, K. Tatsuno, K. Kataoka, and Y. Takeda, "Holographic Video Disk: An Alternative Approach to Optical Video Disks," *Appl. Opt.* **15**, 1398-1403 (1976).
10. I. Satoh and M. Kato, "Holographic Disk Recording of Digital Data with Fringe Stabilization," *Appl. Opt.* **27**, 2987-2992 (1988).
11. T. Yatagai, J. G. Camacho-Basilio, and H. Onda, "Recording of Computer-Generated Holograms on an Optical Disk Master," *Proc. Soc. Photo-Opt. Instrum. Eng.* **1052**, 119-124 (1989).
12. T. Inagaki, "Hologram Lenses Lead to Compact Scanners," *IEEE Spectrum* **26**, 39-43 (1989).
13. B. Brown and A. Lohmann, "Complex Spatial Filtering with Binary Masks," *Appl. Opt.* **5**, 967-969 (1966).
14. W.-H. Lee, "Binary Computer Generated Holograms," *Appl. Opt.* **18**, 3661-3669 (1979).
15. G. Tricoles, "Computer Generated Holograms: An Historical Review," *Appl. Opt.* **26**, 4351-4360 (1987).
16. A. Lohmann, U. Erlangen-Nuremberg; personal communication.
17. J. Alspector and R. B. Allen, "A Neuromorphic VLSI Learning System," *Advanced Research in VLSI Processes 1987 Stanford Research Conference* (MIT Press, Cambridge, 1987), pp. 313-349.
18. Special Issue on Neural Networks, *Applied Optics* **26**, (1 Dec. 1987).
19. J. H. Kim, S. H. Lin, J. Katz, and D. Psaltis, "Monolithically Integrated 2-D Arrays of Optoelectronic Devices for Neural Network Applications," *Proc. Soc. Photo-Opt. Instrum. Eng.* **1043**, 44-52 (1989).
20. A. VanderLugt, "Signal Detection by Complex Spatial Filtering," *IEEE Trans. Inf. Theory* **IT-10**, 139 (1964).
21. D. Psaltis, M. A. Neifeld, and A. A. Yamamura, "Optical Disk Based Correlation Architectures," in *Technical Digest, Topical Meeting on Optical Computing* (Optical Society of America, Washington, DC, 1989), pp. 206-209.
22. D. Psaltis, M. A. Neifeld, and A. A. Yamamura, "Image Correlators Using Optical Memory Disks," *Opt. Lett.* **14**, 429-431 (1989).
23. J. Yu, "Optical Processing Using Photorefractive Crystals," Ph.D. Thesis, California Institute of Technology (1988), Chap. 5.
24. D. Psaltis, "Incoherent Electrooptic Image Correlator," *Opt. Eng.* **23**, 12-15 (1984).
25. A. D. Mikaelyan, A. Vanin, E. D. Gulanyan, and S. Prokopenko, "Holographic Disk for Data Storage," *Sov. J. Quantum Electronics.* **170**(5), 680-687 (1987).