INTRODUCTION

Mass memory systems serve computer needs in both archival and backup needs. There exist numerous applications in both the commercial and military sectors that require data storage with huge capacity, high data rates and fast access. To address such needs 3-D optical memories have been proposed. Since the data are stored in volume, they are capable of much higher storage densities than existing 2-D memory systems. In addition this memory system has the potential for parallel access. Instead of writing or reading a sequence of bits at each time, entire 2-D data pages can be accessed at one go. With advances in the growth and preparation of various photorefractive materials, along with the advances in device technologies such as spatial light modulators(SLM), and detector arrays, the realizations of this optical system is becoming feasible.
HOLOGRAMS

A hologram is a recording of the optical interference pattern that forms at the intersection of two coherent optical beams. Typically, light from a single laser is split into two paths, the signal path and the reference path. The beam that propagates along the signal path carries information, whereas the reference is designed to be simple to reproduce. A common reference beam is a plane wave: a light beam that propagates without converging or diverging. The two paths are overlapped on the holographic medium and the interference pattern between the two beams is recorded. A key property of this interferometric recording is that when it is illuminated by a readout beam, the signal beam is reproduced. In effect, some of the light is diffracted from the readout beam to “reconstruct” a weak copy of the signal beam. If the signal beam was created by reflecting light off a 3D object, then the reconstructed hologram makes the 3D object appear behind the holographic medium. When the hologram is recorded in a thin material, the readout beam can differ from the reference beam used for recording and the scene will still appear.

VOLUME HOLOGRAMS

To make the hologram, the reference and object beams are overlapped in a photosensitive medium, such as a photopolymer or inorganic crystal. The resulting optical interference pattern creates chemical and/or physical changes in the absorption, refractive index or thickness of the storage media, preserving a replica of the illuminating interference pattern. Since this pattern contains information about both the amplitude and the phase of the two light beams, when the recording is illuminated by the readout beam, some of the light is diffracted to “reconstruct” a weak copy of the object beam. If the object beam originally came from a 3–D object, then the reconstructed hologram makes the 3–D
object reappear. Since the diffracted wave front accumulates energy from throughout the thickness of the storage material, a small change in either the wavelength or angle of the readout beam generates enough destructive interference to make the hologram effectively disappear through Bragg selectivity. As the material becomes thicker, accessing a stored volume hologram requires tight tolerances on the stability and repeatability of the wavelength and incidence angle provided by the laser and readout optics. However, destructive interference also opens up a tremendous opportunity: a small storage volume can now store multiple superimposed holograms, each one distributed throughout the entire volume. The destructive interference allows each of these stored holograms to be independently accessed with its original reference beam.

To record a second, angularly multiplexed hologram, for instance, the angle of the reference beam is changed sufficiently so that the reconstruction of the first hologram effectively disappears. The new incidence angle is used to record a second hologram with a new object beam. The two holograms can be independently accessed by changing the readout laser beam angle back and forth.
For a 2-cm hologram thickness, the angular sensitivity is only 0.0015 degrees. Therefore, it becomes possible to store thousands of holograms within the allowable range of reference arm angles (typically 20–30 degrees). The maximum number of holograms stored at a single location to date is 10,000.

**BASIC WORKING.**

In holographic data storage, light from a coherent laser source is split into two beams, signal (data-carrying) and reference beams. Digital data to be stored are "encoded" onto the signal beam via a spatial light modulator. The light of the signal beam traverses through the modulator and is therefore encoded with the "checkerboard" pattern of the data page. This encoded beam then interferes with the reference beam through the volume of a photosensitive recording medium, storing the digital data pages.

The interference pattern induces modulations in the refractive index of the recording material yielding diffractive volume gratings. The reference beam is used during readout to diffract off of the recorded gratings, reconstructing the stored array of bits. The reconstructed array is projected...
onto a pixelated detector that reads the data in parallel. This parallel readout of data provides holography with its fast data transfer rates. Because of the thickness of the hologram, this reference wave is diffracted by the interference patterns in such a fashion that only the desired object beam is significantly reconstructed and imaged on an electronic camera. The theoretical limits for the storage density of this technique are around tens of terabits per cubic centimeter.

A large number of these interference gratings or patterns can be superimposed in the same thick piece of media and can be accessed independently, as long as they are distinguishable by the direction or the spacing of the gratings. This separation is achieved by multiplexing schemes that include changing the storage reference angle, wavelength, or phase code. Among the angle multiplexing is the easiest. In addition to high storage density, holographic data storage promises fast access times, because the laser beams can be moved rapidly without inertia, unlike the actuators in disk drives. With the inherent parallelism of its page wise storage and retrieval, a very large compound data rate can be reached by having a large number of relatively slow, and therefore low-cost, parallel channels.
STORING AND RETRIEVING DIGITAL DATA

The main hardware components are the SLM used to imprint data on the object beam, two lenses for imaging the data onto a matched detector array, a storage material for recording volume holograms, a reference beam intersecting the object beam in the material, the laser source, beam-forming optics for collimating the laser beam, beam splitters for dividing the laser beam into two parts, stages for aligning the SLM and detector array, shutters for blocking the two beams when needed, and waveplates for controlling polarization.

Assuming that holograms will be angle-multiplexed (superimposed yet accessed independently within the same volume by changing the incidence angle of the reference beam), a beam-steering system directs the reference beam to the storage material. Wavelength multiplexing has some advantages over angle-multiplexing, but the fast tunable laser sources at visible wavelengths that would be needed do not yet exist. The optical system with two lenses separated by the sum of their focal lengths, is called the “4-f” configuration, since the SLM and detector array turn out to be four focal lengths apart. Other imaging systems such as the Fresnel configuration (where a single lens satisfies the imaging condition between SLM and detector array) can also be used, but the 4-f system allows the high numerical apertures (large ray angles) needed for high density. In addition, since each lens takes a spatial Fourier transform in two dimensions, the hologram stores the Fourier transform of the SLM data, which is then Fourier transformed again upon readout by the second lens. This has several advantages: Point defects on the storage material do not lead to lost bits, but result in a slight loss in signal-to-noise ratio at all pixels; and the storage material can be removed and replaced in an offset position, yet the data can still be reconstructed correctly. In addition, the Fourier transform properties of the 4-f system lead to the parallel optical search capabilities offered by holographic associative retrieval. The disadvantages of the Fourier transform geometry come from the uneven distribution of intensity in the shared focal plane of the two lenses.
To use volume holography as a storage technology, digital data must be imprinted onto the object beam for recording and then retrieved from the reconstructed object beam during readout. The device for putting data into the system is called a spatial light modulator (SLM)—a planar array of thousands of pixels. Each pixel is an independent microscopic shutter that can either block or pass light using liquid-crystal or micro-mirror technology. Liquid crystal panels with 1024*1024 pixels, and micro-mirror arrays with 1024*768 elements, are commercially available due to the success of computer-driven projection displays. The pixels in both types of devices can be refreshed over 1000 times per second, allowing the holographic storage system to reach an input data rate of 1 Gbit per second—assuming that the laser power and material sensitivities permit.
The data are read using an array of detector pixels, such as a CCD camera or CMOS sensor array. The object beam often passes through a set of lenses that image the SLM pixel pattern onto the output pixel array, as shown. To maximize the storage density, the hologram is usually recorded where the object beam is tightly focused. To access holographically–stored data, the correct reference beam must first be directed to the appropriate spot within the storage media. The hologram is then reconstructed by the reference beam, and a weak copy of the original object beam continues along the imaging path to the camera, where the optical output is detected and converted to digital data. The speed of a storage device can be jointly described by two parameters: the readout rate (in bits per second) and the latency, or time delay between asking for and receiving a particular bit of data. The latency tends to be dominated by mechanical movement, especially if the storage media has to be moved. The readout rate is often dictated by the camera integration time: the reference beam reconstructs a hologram until a sufficient number of photons accumulate to differentiate bright and dark pixels. A frequently mentioned goal is an integration time of about 1 millisecond, which implies that 1000 pages of data can be retrieved per second. If there are 1 million pixels per data page and each pixel stores one bit, then the readout rate is 1 Gigabit per second. This goal requires high laser power (at least 1 W), a storage material capable of high diffraction efficiencies, and a ‘megapixel’ detector (one with a million pixels) that can be read out at high frame rates. Despite these requirements, even faster readout and lower latency could be reached by steering the reference beam angle non–mechanically, by using a pulsed laser, and by electronically reading only the desired portion of the detector array. Both capacity and readout rate are maximized when each detector pixel is matched to a single pixel on the SLM, but for large pixel arrays this requires careful optical design and alignment. In order to study the recording physics, materials, and systems issues of holographic digital data storage in depth, we have built several precision holographic recording testers on which this pixel–to–pixel matching has been achieved.
MULTIPLEXING SCHEMES

Multiple data pages can be stored in a single piece of photorefractive medium via multiplexing. Then they can be retrieved without crosstalk. The three most commonly used schemes are angular multiplexing, wavelength multiplexing, wavelength multiplexing.

In angular multiplexing, the address of each data page is represented by the angle of the incident beam. To change the angle of the beam a mirror mounted on a stepper motor can be used. But this introduces mechanical inertia so to eliminate this acoustooptic deflectors are used (two of them must be used to eliminate Doppler shift in frequency making the system more complex.).

For wavelength multiplexing both the object and reference beams are fixed and only their wavelength is changed. The research on this type of multiplexing is stimulated by the development of the tunable laser diode and photorefractive materials sensitive to the wavelength ranges of them.

In phase code multiplexing, the reference beam consists of multiple plane wavefronts. The relative phase between all these wavefronts are adjustable and represent the address of the images. Each image can be reconstructed by illuminating the holograms with the exact same phase code as for the recording the original message.

Out of the three, phase code multiplexing is the most popular because it eliminates any mechanical movement, provides for fast access, high light efficiency and a fixed wavelength. The number of images to be multiplexed is only limited by the number of pixels on the spatial light modulator.
THE PHOTOREFRACTIVE EFFECT

The photorefractive (PR) effect is consecutive to a non-uniform illumination of the sample that creates an index grating by electro-optic effect. Photorefractivity requires centres photoionization, free carriers transport and trapping with charge separation. The resulting space charge field induces a quasi-permanent index grating. The linear electro-optic effect is of tensorial origin. The electro-optic coefficients relate the coefficients \( \frac{1}{n^2} \) to the internal electric field \( E_1 \) by

\[
\delta \left( \frac{1}{n^2} \right)_i = \sum_j r_{ij} E_j .
\]

For zinc blende structure crystals of point group13 m (such as GaAs and CdTe), there is only one independent component of the electro-optic tensor. The expression of the index variation an becomes therefore simple

\[
\delta n = \frac{1}{2} n_b^3 rE
\]

where \( E \) and \( n_b \) are the space charge field magnitude and the refractive index of the non-excited medium respectively. Notice that the PR index effect is anisotropic. The process of space charge field generation is sketched in the figure I. The sample is first illuminated by photons the energy of which satisfies the relation \( h\nu < E_g < 2 h\nu \), where \( E_g \) is the band gap energy. We consider that carriers are generated by deep centres photoionization. The photogeneration of carriers by two-photon absorption is neglected. For simplicity, we assume the presence of a unique donor centre that may generate only free electrons in the conduction band. At thermal equilibrium, before excitation of the sample, donor centres are partially ionized because of the presence of shallow acceptor centres. This is the so-called compensation mechanism. The interference of two excitation beams induces a non uniform illumination of the crystal the spatial distribution of the crystal illumination is sinusoidal. In-
enlightened regions, electrons are generated and donors are ionized. Because of transport mechanisms, the conduction band electrons migrate towards the dark regions. During this time, the donor charge stays unchanged. After a delay depending on the efficiency of the processes of carrier capture, the electrons recombine on the ionized donors. Thus, there is an excess of positive charges (ionized donors) in the illuminated regions whereas there is an excess of negative charges (by non compensated ionized acceptors) in the dark regions. The spatial variation of electric charge is permanent and induces a space charge field $E_{sc}$ that induces a spatial sinusoidal modulation of the refraction index via the electro-optical effect. Notice that these two gratings (field and index) are phaseshifted of $\pi/2$ compared with the illuminated pattern (cf. Poisson equation). The photoinduced index grating persists as long as there is no thermal reemission of electrons by the donors. So, it persists as long as the space charge field persists. The
processes of photogeneration, transport and capture of the electrons as well as the process of field creation can be incorporated into a set of four equations, namely, two rate equations for the electron density $n$ and the ionized donor density $D^+$, one equation for the current $j_n$ (4c) and one for the electric field $E_{ac}$. For a one dimensional problem, the system reads

\[
\frac{\partial n}{\partial t} = \sigma_n \Phi (D_T - D^+) - c_n D^+ n + \frac{1}{e} \frac{\partial j_n}{\partial x}
\]

\[
\frac{\partial D^+}{\partial t} = \sigma_n \Phi (D_T - D^+) - c_n D^+ n
\]

\[
j_n = e \mu_n E_{ac} n + e D_n \frac{\partial n}{\partial x} \quad e > 0
\]

\[
\frac{\partial E_{ac}}{\partial x} = \frac{e}{\varepsilon} (D^+ - n - A^-)
\]

with the initial conditions

\[
\Phi(x) = \frac{l_p}{h \nu} \left\{ 1 + m \cos \left( \frac{2 \pi}{A} x \right) \right\} \text{ during the pulse}
\]

\[
A^- (t = 0) = D^+ (t = 0)
\]

where $l_p = l_{p1} + l_{p2}$ is the illumination intensity due to the two excitation beams, $m = 2 (l_{p1}, l_{p2})^{0.05}/(l_{p1} + l_{p2})$ is the modulation rate of the interference fringes, $m_n$, $D_n$ are the mobility and diffusion coefficient of the electrons respectively. $D_T$ is the total density of deep donors. $A^-$, $s_n$, $c_n$ are respectively the density of ionized shallow acceptors, the optical cross section for electron emission from neutral donor centres and the electron capture coefficient. The index modulation by photorefractive effect enables to reach diffraction efficiencies larger than $10^{-3}$. A stable index grating appears only after migration and trapping of the free carriers (here the electrons) by the deep centres (donors). This time of grating build-up depends on many crystal parameters such as the carrier photogeneration rate, the mobility and capture coefficients. Depending on the material and on the experimental conditions, the formation time varies from few nanosecond to few seconds.
PHOTOREFRACTIVE MATERIALS

When two coherent plane waves with equal amplitude interfere on this photorefractive medium, the bright regions will be positively charged while the dark regions are negatively charged. The resulting space field changes the photorefractive index of the medium.

Photorefractive effect can be found in many materials, the three most commonly used photorefractive materials in optical memories fall into three categories: electrooptic crystals; semi-insulating compound semiconductors; and photopolymers. Lithium niobate (LiNbO₃), barium titanate (BaTiO₃), and strontium barium niobate (SBN Sr⁽¹₋ₓ⁾BaₓNb₂O₆).

Iron doped lithium niobate has a large index modulation due to photovoltaic effect. It is available in large dimensions and because of it’s strong mechanical qualities it can be used in making holographic discs. high diffraction efficiency is possible with barium titanate. This eliminates the phase distortion of the image beam. But it is not available in large dimensions. SBN has a large electrooptic coefficient (r₃₃₃) which can be reached without any special cut. The crystalline phase distortion, which can be tuned by doping, is far from the room temperature. The photorefractive effect varies between different samples of the same material and from ingot to ingot.

Photopolymer is a new type of photorefractive material. Recently, a photopolymer based on a photoconductor poly(N-vinylcarbazalo) doped with nonlinear optical chromophore was developed and it has better photorefractive properties than inorganic photorefractive crystals.

Photosensitive materials for volume holography are generally classified as either read-write or write-once.

**Read-write materials:** Most holographic read-write materials are inorganic photorefractive crystals doped with transition metals such as iron or rare-earth ions such as praseodymium, grown in large cylinders in the same way as semiconductor materials. Large samples can be cut and polished, making thick holograms possible. These materials react to the light and dark regions of an interference
pattern by transporting and trapping photo-ionized electrons. Through the linear electro-optic effect exhibited by these crystals, the electrical fields created by the trapped charge give rise to an index or phase grating suitable for diffracting light. Thus, the spatial variations in light intensity present in the interference pattern become identical variations in the index of refraction. The trapped charge can be rearranged by subsequent illumination, which makes it possible to erase recorded holograms and replace them with new ones. However, the ease of charge re-excitation also results in the gradual erasure of stored holograms during normal readout. In the dark, the lifetime of these holograms ranges from months to years as the trapped charge slowly leaks away. Recorded holograms can be “fixed” (made semi permanent and resistant to erasure during readout) through thermal or electronic processes. The fixing process affects all the stored holograms within a volume simultaneously. Thus, individual pages of data cannot be erased and replaced this way. An alternative for achieving nonvolatile storage in photorefractive materials is to record at a light wavelength not normally absorbed by the crystal except in the presence of a third “gating” beam of different wavelength. This beam is present only during the recording and is switched off for readout. Organic photorefractive polymers have also been developed. These materials provide more opportunity for performance tuning because you can fabricate them using a wide variety of constituents. However, these materials tend to be limited in thickness and require large applied voltages.

**Write-once materials**: Writing permanent volume holograms generally involves irreversible photochemical reactions, triggered by the bright regions of the optical interference pattern. For example, a photopolymer material will polymerize (bind short monomer chains together to form long molecular chains) in response to optical illumination. In contrast, the molecules in a photochromic material undergo a change in their absorption behavior. Such materials are inexpensive to make in quantity. However, both types can have problems reproducing the object beam faithfully—the
photopolymer because of shrinkage, the photochromic because of oversensitivity to average local intensity. Careful system design can minimize these problems. One advantage of a photopolymer is that after recording, any leftover monomers can be disposed of without affecting the recorded holograms. A photochromic material, however, requires a separate chemical or optical step to disable the unused absorbing molecules after the holograms are recorded. Currently available versions of these write-once materials are thin (approximately 100 um)—the difficulties in making thick samples include insufficient optical quality or excessive absorption. As we will show later, however, new multiplexing techniques for thin materials have made write-once photopolymers one of the leading candidates for the first holographic memory products.

**DYNAMIC RANGE**

In the readout process, the reconstructed hologram is imaged onto the output detector array, where the digital data is extracted from the detected signal. Noise can cause errors to occur during the detection process in various ways. The basic trade-off in volume holography is caused by the fixed noise floor and the finite dynamic range of the recording material. In other words,

- the electronic detection process at the camera contributes the same amount of noise, no matter how bright the hologram, and
- as the number of holograms or the readout rate increases, the amount of power diffracted into each hologram reconstruction decreases. Even if all other noise sources are negligible, there will be a certain hologram strength at which the signal-to-noise ratio is inadequate for error-free detection. The number of detected electrons per pixel can be written as

\[
\eta_{\text{electrons}} \propto \frac{M/\# P_{\text{readout}}}{M^2 N_{\text{pixels}}} \left( \frac{r_{\text{readout}}}{N_{\text{pixels}}} \right),
\]
where $M$ is the number of multiplexed holograms, $N_{\text{pixels}}$ the number of pixels per hologram, $t_{\text{readout}}$ the integration time of the camera, $P_{\text{readout}}$ the power in the readout beam, and $M/\#$ a material/system constant, which measures dynamic range. The storage capacity is $MN_{\text{pixel}}$ and the readout rate is $N_{\text{pixel}}/t_{\text{readout}}$. An increase in either of these parameters leads to a decrease in the number of signal electrons. Given the minimum acceptable number of signal electrons per pixel, we can maximize the capacity and readout rate by increasing $P_{\text{readout}}$ or $M/\#$. Different processes determine the $M/\#$ constant in photorefractives and write-once media. In a photorefractive crystal, the holograms’ recording exposures must be carefully scheduled to record equal-strength holograms. The first hologram is made quite strong. This first hologram erases slowly while the other holograms are stored, and finishes at the same strength as the weakly written final hologram. Alternatively, all the holograms can be cycled several times. The equalized diffraction efficiency falls as one over the square of the number of holograms, with the $M/\#$ as the proportionality coefficient:

$$n = \left( \frac{M/\#}{M} \right)^2$$

The $M/\#$ constant in a photorefractive material becomes large if the holograms can be recorded faster than they erase. In iron-doped lithium niobate, a typical $M/\#$ might be 1. This implies that to store 1,000 holograms with 1 million pixels and read each in 1 millisecond, we need about 1W in the reference beam. A write-once material has much in common with photographic film: After a finite amount of input energy, the material is completely exposed. Each hologram gets its share of the dynamic range as it is recorded, preserving the bright and dark regions of the interference fringes. For instance, in a photopolymer material, the photosensitivity saturates as the available supply of monomers is exhausted. It turns out that the diffraction efficiency of individual holograms, when $M$ of them are multiplexed in a saturable medium such as a photopolymer, also follows the $(M/\#/M)^2$ relationship. The most commonly used polymer is DuPont’s HRF-150. The 100-micron-thick version has a $M/\#$ of 6.5$^2$, which reduces the required readout power by a factor of 40.
UNDERSTANDING NOISE

In addition to detector noise, other factors can cause errors:

- **The readout conditions change.** This can occur, for instance, when the recording alters the recording material properties. This causes unwanted changes in the reference beam path between the time the hologram is recorded and the time it is reconstructed. Often, the reference beam angle or wavelength can be tuned to optimize the diffraction efficiency and partially compensate for this effect.

- **The detector array doesn’t line up with the pixel array in the reconstructed hologram.** This includes errors in camera registration, rotation, focus, tilt, and image magnification.

- **The detector is receiving undesired light,** either from light scattering off the storage material, crosstalk from other stored holograms, or crosstalk between neighboring pixels of the same hologram. Although crosstalk contributions scale with the strength of the holograms, scattering depends only on readout power and the components’ optical quality.

- **There are brightness variations across the detected image.** This can be a problem if a single threshold is used across the image to separate the pixels into bright and dark and to assign binary values. These fluctuations can be caused by the SLM, the optical imaging, or the original laser beams.

COMBATING ERRORS

Commercial storage products have user error rates as low as $10^{-15}$. For a 1-Gbyte hard drive, a user might expect to read the entire drive 100,000 times before a single bit error. On the other hand, at the hardware level, errors occur at a rate of perhaps 1 in 10,000 bits read. The designers decrease the error rate that the user sees by storing redundant its along with user data. The sequence of user and
redundant bits forms an error correction code. ECC algorithms performed in hardware after the read head can detect if a few bits within each code word are in error and then pass on corrected user bits. The redundant bits cause a slight sacrifice in the capacity of each individual hologram. This sacrifice is measured by the ECC code rate—the fraction of bits stored that are actual user data. However, the increase in the raw bit error rate (BER) that can be tolerated allows additional pages to be stored, which increases overall capacity. A typical ECC code with a code rate of 0.9 can handle an input data stream with a raw BER of $10^{-4}$. It can output the user’s data, stripped of redundant bits, with the desired BER of $10^{-15}$. Since the ECC code rate drops quickly as the expected raw BER climbs above $10^{-3}$, most holographic memory designs aim for a raw BER of $10^{-4}$ or so. Keeping the raw BER at this target value takes careful engineering of the optical system combined with signal processing and modulation coding. Careful engineering of the optics alone is not generally the most cost-effective solution to reducing the BER to the target value.

**SIGNAL PROCESSING**

Signal processing works by considering the storage device as an imperfect transmission “channel” for data that tends to smear together the signal energy from multiple bits of user data. Knowledge of how this intermixing occurs can be applied at the output end to eliminate the crosstalk and reproduce the originally transmitted bit sequence. In a telecommunications application or a bit-serial storage device like a hard drive or DVD disk, the smearing takes place between signals adjacent in time. In holographic storage, the smearing occurs spatially in two dimensions, as light intended for a particular CCD pixel diffracts into neighboring pixels. Signal processing techniques for holographic storage are therefore 2D extensions of the 1D techniques developed for bit-serial devices. Examples of signal processing techniques used in holographic memories are adaptive thresholding and normalization, equalization, filtering, and partial response precoding at the input.
MODULATION CODES

A modulation code dictates the way in which bits of information are encoded into the channel as data signals. They are selected to facilitate the detection process and hence improve overall performance. For instance, in bit-serial devices, modulation codes are used to set upper and lower bounds on the frequency at which the signal level changes. In holographic storage, modulation codes are used to avoid pixel combinations that are prone to distortion and to create easy-to-detect pixel patterns. A convenient encoding that facilitates detection is the organization of binary data into small blocks of pixels, such that the number of bright pixels is constant (usually half the pixels). The simplest example is differential encoding, in which 2 pixels convey 1 bit of information. This technique was used for storing digital data by the group at Stanford University. Several modulation codes with higher code rate and performance, and thus higher complexity, have since been developed for holographic storage. These codes have been used to demonstrate as many as 1,200 superimposed holograms in lithium niobate (LiNbO₃) at a raw BER of $10^{-8.9}$
SPATIAL MULTIPLEXING

In most cases, we must increase the capacity to at least 1 terabit to have a system that is competitive with alternative technologies. We can accomplish this by constructing a large memory consisting of multiple 1-Gbit modules. This technique is called spatial multiplexing, because multiple “stacks” of holograms are stored in different spatial locations of the recording material. Spatial multiplexing has several configuration options: holographic random-access memory (HRAM), compact modular holographic memory, and holographic 3D disks.

1. HOLOGRAPHIC RANDOM-ACCESS MEMORY

One approach for spatial multiplexing steers the reference and object beams to a stationary block of material containing multiple storage locations, as Figure 3 shows.1,7 With nonmechanical optical scanners, the HRAM system can very rapidly steer the optical beams. Most nonmechanical beam steerers use either an acousto-optic deflector or a one-dimensional liquid crystal SLM. By using large lenses (not shown in Figure 3), the information stored at separate locations can be directed back to a single detector array. An HRAM system can read out holograms from any location in an essentially random sequence. To maximize the number of holograms in each location, designers generally envision HRAM systems with thick read-write materials such as photorefractive crystals. Researchers at Caltech built a 16-location HRAM system capable of 10,000 holograms per location1,7; researchers at Rockwell demonstrated an HRAM system with no moving parts.2 To construct a Tbit memory using this approach, we need 1,000 spatial locations (arranged in 2D as a 33 × 33 array), with each location storing 1 Gbit. The main challenge in building such a system is the optics that have to simultaneously transfer data from each of the 1,000 recording sites on the recording material to a single detector array. This will require considerable
engineering improvements over present systems. 2,7

**Recording rate.** The photosensitivity of most photoreflective crystals is relatively low; therefore, the recording rate is invariably one to two orders of magnitude slower than the readout rate. In addition, it is practically impossible to change the state of a single pixel within a stored hologram, and it is possible but not easy to replace a single hologram within a hologram stack.4 Instead, an entire stack of holograms must be erased together, either by heating or by illumination with the “gating” light. Therefore, an HRAM system is not truly a read-write memory; more accurately, it is an erasable write-once, read-many memory.

**Readout rate.** The readout rate in an HRAM system is mostly limited by the camera integration time. The reference beam reconstructs the same hologram until a sufficient number of photons accumulate to differentiate bright and dark pixels. An oft-mentioned goal is an integration time of about 1 millisecond, leading to 1,000 data pages per second. If there are 1 million pixels per data page, then the readout rate is 1 gigapixel per second. In the HRAM system this data rate can be supported continuously, independent of readout order. The latency is dominated by the integration time and is typically about 1 millisecond. This can be reduced by using a pulsed laser or a CMOS detector array or both.

**Applications.** The HRAM system is best matched to applications with high capacity and fast readout rate demands, yet with relatively infrequent changes to the stored data. Video-on-demand and Web servers fit well here: movie and Web content change infrequently, yet multiple users are continuously accessing enormous amounts of content in a fairly random order. (Playing one movie for a single user is sequential— playing 10 movies at once is not.) An alternative method for reaching 1 Tbit in storage capacity is a jukebox-type apparatus. Blocks of material, each containing 1 Gbit or more, are brought into position in front of the reference beam optics for readout. The access time to a hologram is either 1 millisecond if the hologram is in the current material block,
or several seconds if it is in a separate block. This can be reduced somewhat by having several readout stations. A principal advantage of increasing the capacity in this way is in the cost per megabyte of storage. For a one-block HRAM system, the cost is dominated by the components: camera, SLM, laser, beam steerers, and optics. The advantages provided by the lack of moving parts are probably enough to support this cost per megabyte only for military applications. For the commercial market, however, the cost per megabyte drops rapidly as more blocks are used, until the cost of the material becomes dominant.

**Associative retrieval.** An HRAM system also lets designers use a unique feature of holographic storage: associative retrieval. To search a conventional storage device for all data records sharing a particular feature, we would retrieve each record into RAM, search it using software, and continue until all records were recalled and checked. With holographic storage, this process can be performed at the memory itself. Instead of reconstructing signal data pages with a reference beam, the data pattern of interest is put on the SLM and illuminates the storage location with the signal beam. All the reference beams used to store holograms in that stack are reconstructed. The brightness of each beam, however, is proportional to the correlation between the original stored data pattern and the data pattern from the interrogating signal beam. Once the reference beams are focused onto a “correlation plane” detector array, the reference beam angles corresponding to the closest matches can be identified. The reference beam can then be used to reconstruct the desired data page onto the output camera, completing the search-and-retrieval process in perhaps 5 milliseconds.

### 2. COMPACT MODULAR HOLOGRAPHIC MEMORY

One drawback in the HRAM system is that the number of rapidly accessible locations (and thus the immediately accessible capacity) is limited by the beam-steering optics. Rather than bring the beams to the storage material, another approach is to bring the pixel arrays for data input and output to
the storage material. In fact, by applying a unique feature of the stored holograms, the same pixel array can be used for both input and output. Upon readout, instead of bringing back the same reference beam used during recording, its “phase conjugate” is directed to the storage location. (The phase conjugate of an optical beam passes backward along the beam’s path, like a movie played in reverse.) This new readout beam reconstructs the phase conjugate of the signal beam, which returns along the original signal path back to the SLM. Because of this, a phase conjugate signal beam allows the use of a cheap imaging lens, or even no lens at all. If each pixel of the SLM is not only a light modulator but also a detector, then the entire storage device can be fabricated from identical compact modules with no moving parts. One such module is shown in Figure 4. With several modules, the memory resembles a board of DRAM, with holography increasing the amount of data stored per RAM chip. Caltech researchers recently demonstrated a single-module compact holographic system with 480 modulator/detector pixels and 25 stored holograms. A small amount of logic at each “smart” pixel let the system counteract the erasure in a photorefractive crystal by periodically detecting and refreshing the holograms. This can remove the need to fix the stored holograms and makes it feasible to individually erase holograms from a stack. In a modular system, the cost per megabyte is dominated by the smart pixel array and the two compact angle steerers (one for the writing beam, one for the readout). Associative retrieval is possible, but adds another detector array per module. Increasing the number of pixels while keeping the cost of the detector array and angle tuners low is the key to practical implementation of this architecture.

3. HOLOGRAPHIC 3D DISKS

A third approach to spatial multiplexing leaves the optics and components stationary and moves the storage material. The simplest method to do this employs a disk configuration. The disk is constructed with a thick layer (approximately 1 mm) of the holographic material. Multiple holograms can be stored at each location on the disk surface. These locations are arranged along radial tracks,
with the motion of the head selecting a track, and the disk rotation providing access along each track. As the medium thickness increases, the number of holograms that can be stored increases (and therefore the surface density goes up). However, the surface area that is illuminated also increases (and therefore the surface density goes down). Typically, it would be desirable to fabricate disks of 1-mm thickness, yielding a density of approximately 100 bits per squared micron. Even though the data is still stored in 3D, in the disk configuration the surface density, not the volume density, is what matters for most practical purposes. Angle multiplexing can be used to multiplex holograms on the holographic disk in a manner similar to the HRAM architecture. However, the angle scanner would make the readout head too large and heavy for rapid access to holograms on different radial tracks. A single, simple reference beam that could attain the same density without a bulky beam deflector would be more convenient. This can be done by making the reference beam a spherical or converging beam. This is effectively equivalent to bringing in all the reference beam angles simultaneously instead of one at a time. In this case the reconstruction becomes very sensitive to the recorded hologram’s position instead of the reference’s angle of incidence. In fact, if the material is shifted by a few microns, the reconstruction disappears and a second hologram can be stored. The motion of the material relative to the illuminating spherical beam needed to reconstruct different shift-multiplexed holograms is conveniently supplied by the disk’s rotation. In addition, the simplicity of the reference beam makes the readout system look like a CD/DVD disk, albeit with parallel readout, as shown in Figure 5. The shift-multiplexing technique also depends on material thickness. Experimentally, it has been demonstrated that a thickness of 1 mm is sufficient to support 100 bits/µm², 20 times the areal density of a single-layer DVD. Holographic disks can be configured as either a WORM or a ROM system. A WORM system incorporates an SLM, turning the read head into a readwrite head. The recording procedure is complicated by the chemical reactions in the photopolymers that are the recording material for the 3D disk. These reactions are not driven by light so much as triggered by it.
Once begun (within an illuminated region), the reaction continues after the optical exposure stops. Because each hologram position in a shift-multiplexed WORM disk overlaps many others, the entire disk has to be recorded without stopping in order to reach the maximum capacity. ROM applications, where the user buys a written disk (movies, audio, a computer game) and owns a simple read-only unit, may be best suited for the shift-multiplexed disk. The ROM system requires a master disk that is very similar to the WORM disk. Since the mastering device is not sold to users, it can be a large, expensive apparatus. The master disk is copied by bringing a blank disk in contact with the master and illuminating the two disks together. Holographic disks can be viewed as a candidate technology to succeed the recently introduced DVD. They offer both higher storage density and speed. The primary factor preventing commercialization of holographic disks is the lack of a photopolymer material with sufficient thickness. Currently available materials have a thickness of 100 microns, yielding a surface density of only 10–20 bits per squared micron, which is the same order as the DVD. Research in material development promises to increase the thickness to 1 mm, increasing the achievable storage density by an order of magnitude.

**PRODUCT SCENARIOS**

Holographic storage is a promising candidate for next-generation storage. Recent research has demonstrated that holographic storage systems with desirable properties can be engineered. The next step is to build these systems at costs competitive with those of existing technologies and to optimize the storage media. If suitable recording materials become available from the research efforts currently under way, we envision a significant role for holographic storage. Three specific configurations illustrate possible examples of future holographic memories:
1. **Erasable write-once, read-many drives** supporting terabytes of storage, 1 Gbit/second readout rates, and fast access to data in blocks of 50–100 Gbytes. Suitable applications include video on demand and large Web servers.

2. **Write-once 3D disks** supporting more than 100 Gbytes per 120-mm disk. Access time to 100-Mbyte blocks of 10–100 milliseconds, with readout rates of more than 500 Mbits/sec. Suitable applications include archiving of data requiring permanent storage yet rapid access, such as medical data and high-resolution maps and satellite images.

3. **Pre-recorded 3D disks** supporting more than 100 Gbytes per 120-mm disk and readout rates greater than 200 Mbits/sec. Suitable applications include distributing computer programs, movies, and multimedia.
CONCLUSION

Optical and electronic computers will soon require memories with capacities beyond those of magnetic or electronic systems. 3-D storage and parallel access are needed for densities greater than 1TBit/cm³, and I/O bandwidth in excess of 1 GBit/s. Large 3-D memories offer a promise to meet such requirements. With state of the art yet off-the-shelf optoelectronic devices, (SLM’s, CCD array detectors, optical array generators etc.) thousands of data pages can be stored in the same volume of photorefractive crystal with phase code and rotation multiplexing. This approaches the theoretical limitation of a 3-D storage density.

A fast, compact, rugged, low-cost, and very high density 3-D holographic memory. If developed will have a significant impact on today’s data storage community and will have wide applications in areas of supercomputing, information highway, virtual reality, artificial intelligence, fingerprint analysis, medical diagnostics, large scale databases, etc.
REFERENCES


3. Bell laboratories physical sciences research, “High density Holographic data storage”.


5. Yoshiko Hara, “Sony spin-off prototypes Tbyte holographic storage “, EE Times

6. Demetri Psaltis California Institute of Technology Geoffrey W. Burr IBM Almaden Research Center “Holographic Data Storage”