INTRODUCTION

As processors and buses roughly double their data capacity every three years (Moore's Law), data storage has struggled to close the gap. CPUs can perform an instruction execution every nanosecond, which is six orders of magnitude faster than a single magnetic disk access. Much research has gone into finding hardware and software solutions to closing the time gap between CPUs and data storage. Some of these advances include cache, pipelining, optimizing compilers, and RAM. As the computer evolves, so do the applications that computers are used for. Recently large binary files containing sound or image data have become commonplace, greatly increasing the need for high capacity data storage and data access. A new high capacity form of data storage must be developed to handle these large files quickly and efficiently. Holographic memory is a promising technology for data storage because it is a true three dimensional storage system, data can be accessed an entire page at a time instead of sequentially, and there are very few moving parts so that the limitations of mechanical motion are minimized. Holographic memory uses a photosensitive material to record interference patterns of a reference beam and a signal beam of coherent light, where the signal beam is reflected off of an object or it contains data in the form of light and dark areas. The nature of the photosensitive material is such that the recorded interference pattern can be reproduced by applying a beam of light to the material that is identical to the reference beam. The resulting light that is transmitted through the medium will take on the recorded interference pattern and will be collected on a laser detector array that encompasses the entire surface of the holographic medium. Many holograms can be recorded in the same space by changing the angle or the wavelength of the incident light. An entire page of data is accessed in this way.

The three features of holographic memory that make it an attractive candidate to replace magnetic storage devices are redundancy of stored data, parallelism, and multiplexing. Stored data is redundant because of the nature of the interference pattern between the reference and signal beams that is imprinted into the holographic medium. Since the interference pattern is a plane wave front, the stored pattern is propagated throughout the entire volume of the holographic medium, repeating at intervals. The data can be corrupted to a certain level before information is lost so this is a very safe method of data storage. Also, the effect of lost data is to lower the signal to noise ratio so that the amount of data that can be safely lost is dependent on the desired signal to noise ratio. Stored holograms are massively parallel because the data is recorded as an optical wave front that is retrieved as a single page in one access. Since light is used to retrieve data and there are no moving parts in the detector array, data access time is on the order of 10 ms and data transfer rate approaches 1.0 GB/sec. Multiplexing allows many different patterns to be stored in the same crystal volume simply by changing the angle at which the reference beam records the hologram.

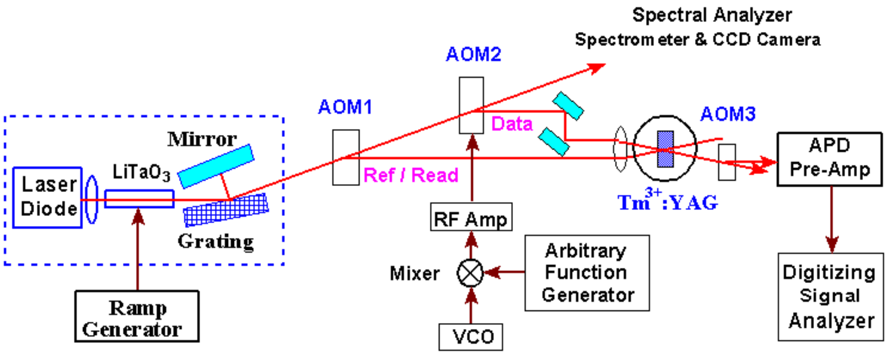
Currently, holographic memory techniques are very close to becoming technologically and economically feasible. The major obstacles to implementing holographic data storage are recording rate, pixel sizes, laser output power, degradation of holograms during access, temporal decay of holograms, and sensitivity of recording materials. An angle multiplexed holographic data storage system using a photorefractive crystal for a recording medium can provide an access speed of 2.4 μs, a recording rate of 31 kB/s and a readout rate of 10 GB/s, which is between the typical values for DRAM and magnetic disk. At an estimated cost of between $161 and $236 for a complete holographic memory system, this may become a feasible alternative to magnetic disk in the near future.

WORKING

A holographic data storage system consists of a recording medium, an optical recording system, and a photodetector array. A beam of coherent light is split into a reference beam and a signal beam which are used to record a hologram into the recording medium. The recording medium is usually a photorefractive crystal such as $LiNbO_3$ or $BaTiO_3$ that has certain optical characteristics. These characteristics are high diffraction efficiency, high resolution, permanent storage until erasure, and fast erasure on the application of external stimulus such as UV light. A ‘hologram’ is simply the three-dimensional interference pattern of the intersection of the reference and signal beams at 90° to each other. This interference pattern is imprinted into the crystal as regions of positive and negative charge. To retrieve the stored hologram, a beam of light that has the same wavelength and angle of incidence as the reference beam is sent into the crystal and the resulting diffraction pattern is used to reconstruct the pattern of the signal beam. Many different holograms may be stored in the same crystal volume by changing the angle of incidence of the reference beam. One characteristic of the recording medium that limits the usefulness of holographic storage is the property that every time the crystal is read with the reference beam, the stored hologram at that "location" is disturbed by the reference beam and some of the data integrity is lost. With current technology, recorded holograms in Fe- and Tb- doped $LiNbO_3$ that use UV light to activate the Tb atoms can be preserved without significant decay for two years.

A series of spectral memory demonstration experiments have been conducted at the University of Oregon. These experiments employ a 780-nm commercial semiconductor diode laser as the light source, a crystal of Tm^{3+}:YAG as the frequency-selective recording material, and an avalanche photodiode as a signal detector. The diode laser was stabilized to an external cavity containing a grating and an electrooptic crystal. The intracavity electrooptic crystal provides for microsecond-time-scale sweeping of the laser frequency over roughly one gigahertz. Two storage (reference and data) beams and one reading beam, are created from the output of the single laser source using the beam splitter and the acousto-optic modulators shown in figure.

The beams are focused to a 150 m^2 spot in a Tm^{3+}:YAG crystal. The reference and data beams are simultaneous as are the read and signal beams.



The most common holographic recording system uses laser light, a beam splitter to divide the laser light into a reference beam and a signal beam, various lenses and mirrors to redirect the light, a photorefractive crystal, and an array of photodetectors around the crystal to receive the holographic data. To record a hologram, a beam of laser light is split into two beams by a mirror. These two beams then become the reference and the signal beams. The signal beam interacts with an object and the light that is reflected by the object intersects the reference beam at right angles. The resulting interference pattern contains all the information necessary to recreate the image of the object after suitable processing. The interference pattern is recorded onto the photoreactive material and may be retrieved at a later time by using a beam that is identical to the reference beam (including the wavelength and the angle of incidence into the photoreactive material). This is possible because the hologram has the property that if it is illuminated by either of the beams used to record it, the hologram causes light to be diffracted in the direction of the second beam that was used to record it, thereby recreating the reflected image of the object if the reference beam was used to illuminate the hologram. So, the reflected image must be transformed into a real image with mirrors and lenses that can be sent to the laser detector array.

There are many different volume holographic techniques that are being researched. The most promising techniques are angle-multiplexed, wavelength-multiplexed, spectral, and phase-conjugate holography. Angle- and wavelength- multiplexed holographic methods are very similar, with the only difference being the way data is stored and retrieved, either multiplexed with different angles of incidence of the reference beam, or with different wavelengths of the reference beam. Spectral holography combines the basic principles of volume holography using a photorefractive crystal with a time sequencing scheme to partition holograms into their own subvolume of the crystal using the collision of ultrashort laser pulses to differentiate between the image and the time-delayed reference beam. Phase-conjugate holography is a technique to reduce the total volume of the system (the system includes recording devices, storage medium, and detector array) by eliminating the need for the optical parts between the spatial light modulator (SLM) and the detector. The SLM is an optical device that is used to convert the real image into a single beam of light that will intersect with the reference beam during recording. Phase-conjugate holography eliminates these optical parts by replacing the reference beam that is used to read the hologram with a conjugate reference beam that propagates in the opposite direction as the beam used for recording. The signal diffracted by the hologram being accessed is sent back along the path from which it came, and is refocused onto the SLM, which now serves as both the SLM and the detector.

There are two main classes of materials used for the holographic storage medium. These are photorefractive crystals and photopolymers (organic films). The most commonly used photorefractive crystals used are $LiNbO_3$ and $BaTiO_3$. During hologram recording, the refractive index of the crystal is changed by migration of electron charge in response to the imprinted three-dimensional interference pattern of the reference and signal beams. As more and more holograms are superimposed into the crystal, the more decay of the holograms occurs due to interference from the superimposed holograms. Also, holograms are degraded every time they are read out because the reference beam used to read out the hologram alters the refractive nature of the crystal in that region. Photorefractive crystals are suitable for random access memory with periodic refreshing of data, and can be erased and written to many times. Photopolymers have been developed that can also be used as a holographic storage medium. Typically the thickness of photopolymers is much less than the thickness of photorefractive crystals because the photopolymers are limited by mechanical stability and optical quality. An example of a photopolymer is DuPont's HRF-150. This film can achieve 12 bits/μm^2 with a 100 μm thickness, which is greater than DVD-ROM by a factor of two. When a hologram is recorded, the interference pattern is imprinted into the photopolymer by inducing photochemical changes in the film. The refractive index modulation is changed by changing the density of exposed areas of the film. Stored holograms are permanent and do not degrade over time or by readout of the hologram, so photopolymers are suited for read-only memory (ROM).

SPECTRAL HOLOGRAPHIC MEMORY

Over the last decades, the speed and capacity of magnetic and optical storage devices have increased enormously. Remarkably, the increases have accrued primarily through gradual refinements rather than fundamental technological changes. Now, armed with a new spectral holographic recording technique and a spectrally selective storage material, researchers at the University of Oregon have pushed data storage densities and density bandwidth products to new levels.

The spectral holographic technique employs purely optical addressing to decrease staorage spot size and thereby increase areal density. The minimal spot sizes are set by diffraction and are clearly identifiable. Increases in areal density beyond the diffraction limit is possible only by the introduction of a non-spatial location. Laser frequency constitutes an obvious possibility as a non-spatial addressing parameter .In a memory implemented with frequency used as an addressing parameter, storage locations become addressable through combined spectral and spatial coordinates. Whether one can actually utilize frequency as an addressing parameter depends on the existence of recording materials that respond independently at some distinct frequencies.

Materials are characterized by two frequency scales – the overall absorption bandwidth and the minimum frequency change to which the material is sensitive. The latter quantity represents the minimum spectral channel width that can be employed. The ratio of the overall absorption bandwidth to the minimum spectral channel width tells us the maximum number of spectral channels supported by the specific material. In some materials, millions of distinct spectral channels are available at low temperatures. A spectral memory implemented with 10^6 spectral channels has been calculated to offer areal data densities of more than 10^12 bits/sq in; more than three orders of magnitude higher than possible in a conventional diffraction limited optical memory. Ultimately, the storage density of a spectral memory is limited by the number of atoms available within each spatial-spectral storage location. Analysis indicates that only about 10^4 absorber atoms are needed to record each bit. This is far fewer than the necessary number of atoms per bit needed in conventional memories.

A result of the time-frequency relation is that bits within an optical data stream occupy wider spectral intervals as the data bandwidth increases. Thus if each bit is to be stored in a frequency dimension, the spectral channel width allocated must be increased as the data rate increases. Spectral holographic principles provide mechanisms for sidestepping time-frequency constraints on spectral data density and data bandwidth. Bits do not have to be localized and time-frequency constraints do not apply.

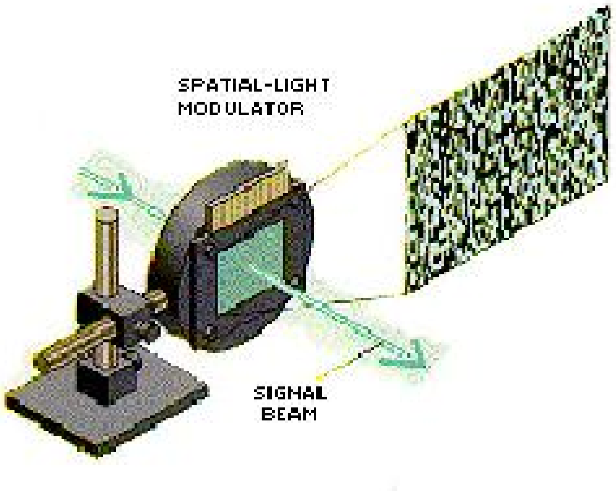
In ordinary spatial holography, interference between two light fields can store laser beam's wave front information. The stored information allows for a beam's complete reproduction. In the newly developed technology of spectral holography, two finite duration beams (simultaneous or not) interact with a frequency selective recording material. Interference of the two beams in frequency space leads to the storage of one beam's temporal waveform information. If the optical beam is encoded with data, that information is included in the recorded waveform. Readout of spectral holograms produces a signal beam whose temporal profile duplicates the original input data beam. Since frequency-selective storage materials are also spatially selective, it is possible to make spatial –spectra holograms in which both the temporal and spatial structure of input beams are recorded.

APPLICATION TO BINARY

In order for holographic technology to be applied to computer systems, it must store data in a form that a computer can recognize. In current computer systems, this form is binary. For this the source beam is manipulated. In computer applications, this manipulation is in the form of bits. The next section explains the spatial light modulator, a device that converts laser light into binary data.

Spatial Light Modulator (SLM)

A spatial light modulator is used for creating binary information out of laser light. The SLM is a 2D plane, consisting of pixels which can be turned on and off to create binary 1's and 0's. An illustration of this is a window and a window shade. It is possible to pull the shade down over a window to block incoming sunlight. If sunlight is desired again, the shade can be raised. A spatial light modulator contains a two-dimensional array of windows which are only microns wide. These windows block some parts of the incoming laser light and let other parts go through. The resulting cross section of the laser beam is a two dimensional array of binary data, exactly the same as what was represented in the SLM. After the laser beam is manipulated, it is sent into the hologram to be recorded. This data is written into the hologram as page form. It is called this due to its representation as a two dimensional plane, or page, of data. Figure below shows a Spatial Light Modulator implemented with a LCD panel.



Page Data Access

Because data is stored as page data in a hologram, the retrieval of this data must also be in this form. Page data access is the method of reading stored data in sheets, not serially as in conventional storage systems. Conventional storage was reaching its fundamental limits. One such limit is the way data is read in streams. Holographic memory reads data in the form of pages instead. For example, if a stream of 32 bits is sent to a processing unit by a conventional read head, a holographic memory system would in turn send 32 x 32 bits, or 1024 bits due to its added dimension. This provides very fast access times in volumes far greater than serial access methods. The volume could be one Megabit per page using a SLM resolution of 1024 x 1024 bits at 15-20 microns per pixel

MULTIPLEXING

Once one can store a page of bits in a hologram, an interface to a computer can be made. The problem arises, however, that storing only one page of bits is not beneficial. Fortunately, the properties of holograms provide a unique solution to this dilemma. Unlike magnetic storage mechanisms which store data on their surface, holographic memories store information throughout their whole volume. After a page of data is recorded in the hologram, a small modification to the source beam before it reenters the hologram will record another page of data in the same volume. This method of storing multiple pages of data in the hologram is called multiplexing. The thicker the volume becomes, the smaller the modifications to the source beam can be.

Angular Multiplexing

When a reference beam recreates the source beam, it needs to be at the same angle it was during recording. A very small alteration in this angle will make the regenerated source beam disappear. Harnessing this property, angular multiplexing changes the angle of the source beam by very minuscule amounts after each page of data is recorded. Depending on the sensitivity of the recording material, thousands of pages of data can be stored in the same hologram, at the same point of laser beam entry. Staying away from conventional data access systems which move mechanical matter to obtain data, the angle of entry on the source beam can be deflected by high-frequency sound waves in solids. The elimination of mechanical access methods reduces access times from milliseconds to microseconds.

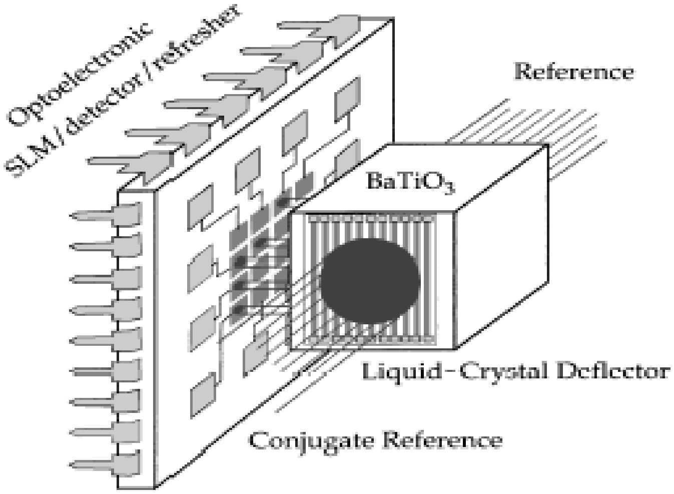


Figure above shows a compact module that uses angular multiplexing. The module is composed of a photorefractive crystal in which holograms are stored, a pair of liquidcrystal beam steerers (one of which is hidden behind the crystal) that is responsible for angularly multiplexing holograms in the crystal, and an OptoElectronic Integrated Circuit (OEIC) that merges the functions of a reflective spatial light modulator (SLM) for recording holograms and a detector array for readout. One is aligned at unit magnification with the photodetectors that sense it, because of the conjugate nature of the readout process and because the detectors are located within the same OEIC pixels as the modulators used to record the holograms. Furthermore, the OEIC provides a solution to the volatility of holograms stored in a read–write photorefractive memory.

Wavelength Multiplexing

Used mainly in conjunction with other multiplexing methods, wavelength multiplexing alters the wavelength of source and reference beams between recordings. Sending beams to the same point of origin in the recording medium at different wavelengths allows multiple pages of data to be recorded. Due to the small tuning range of lasers, however, this form of multiplexing is limited on its own.

Spatial Multiplexing

Spatial multiplexing is the method of changing the point of entry of source and reference beams into the recording medium. This form tends to break away from the non-mechanical paradigm because either the medium or recording beams must be physically moved. Like wavelength multiplexing, this is combined with other forms of multiplexing to maximize the amount of data stored in the holographic volume. Two commonly used forms of spatial multiplexing are peristrophic multiplexing and shift multiplexing.

Phase-Encoded Multiplexing

The form of multiplexing farthest away from using mechanical means to record many pages in the same volume of a holograph is called phase-encoded multiplexing. Rather than manipulate the angle of entry of a laser beam or rotate/translate the recording medium, phase-encoded multiplexing changes the phase of individual parts of a reference beam. The main reference beam is split up into many smaller partial beams which cover the same area as the original reference beam. These smaller beamlets vary by phase which changes the state of the reference beam as a whole. The reference beams intersects the source beam and records the diffraction relative to the different phases of the beamlets. The phase of the beamlets can be changed by non-mechanical means, therefore speeding up access times.

Combining Multiplexing Methods

No single multiplexing method by itself is the best way to pack a hologram full of information. The true power of multiplexing is brought out in the combination of one or more methods. Hybrid wavelength and angular multiplexing systems have been tested and the results are promising. Recent tests have also been formed on spatial multiplexing methods which create a hologram the size of a compact disc, but which hold 500 times more data

ERROR CORRECTION

It is inevitable that storing massive amounts of data in a small volume will be error prone. Factors exist in both the recording and retrieval of information which will be covered in the following subsections, respectively. In order for holographic memory systems to be practical in next generation computer systems, a reliable form of error control needs to be created.

Recording Errors

When data is recorded in holographic medium, certain factors can lead to erroneously recorded data. One major factor is the electronic noise generated by laser beams. When a laser beam is split up (for example, through a SLM), the generated light bleeds into places where light was meant to be blocked out. Areas where zero light is desired might have minuscule amounts of laser light present, which mutates its bit representation. For example, if too much light gets recorded into this zero area representing a binary 0, an erroneous change to a binary 1 might occur. Changes in both the quality of the laser beam and recording material are being researched, but these improvements must take into consideration the cost-effectiveness of a holographic memory system. These limitations to current laser beam and photosensitive technology are some of the main factors for the delay of practical holographic memory systems.

Page-Level Parity Bits

Once error-free data is recorded into a hologram, methods which read data back out of it need to be error free as well. Data in page format requires a new way to provide error control. Current error control methods concentrate on a stream of bits. Because page data is in the form of a two dimensional array, error correction needs to take into account the extra dimension of bits. When a page of data is written to the holographic media, the page is separated into smaller two-dimensional arrays. These sub-sections are appended with an additional row and column of bits. The added bits calculate the parity of each row and column of data. An odd number of bits in a row or column create a parity bit of 1 and an even number of bits creates a 0. A parity bit where the row and column meet is also created which is called an overall parity bit. The sub-sections are rejoined and sent to the holographic medium for recording.

When data is read back from storage, another row and column are added called parity check bits. Because the row of parity bits evens out the data, the addition or subtraction of a bit of stored data will cause two of the parity check bits to become a one. The overall parity check bit becomes a one and the place of error is calculated. The calculation occurs by finding where the column parity check bit and the row parity check bit meet up in the original data. This erroneous bit is flipped and the data is read out error free. If there happens to be two or more errors in the original data, the overall parity check bit becomes a zero and the page is re-read.

INTERFACING

Like error control, the I/O interface to modern computer systems needs to be tailored to data retrieval in page format. Bits are no longer read from a stream, they are sent to the computer as sheets. Clearly the I/O interface needs to be changed to accommodate for this. One of the problems with such large amounts of data being fed to a processor is that the incoming data may exceed the processor’s throughput. This is where interfacing needs to bridge the data in a coherent fashion between memory and processor. In the following subsections, two kinds of interfacing are covered which vary in a unique way.

Smart Interfacing

Smart interfacing is a method of controlling the way data is sent to the processor from holographic memory by a pre-defined set of logical commands. These logical commands come from outside the stored memory and are provided to control the way data is managed before going to the processor. An example of these pre-defined instructions are the fixed set of rules used by error detection and correction. Because these rules stay the same throughout memory retrieval, they can be hard coded into the smart interfacing agent.

Intelligent Interfacing

Seemingly the same as smart interfacing by name,intelligent interfacing is different in one important way. Intelligent interfacing has external control signals which can be manipulated to transform incoming data in a non-static manner. These signals create a way for the intelligent interfacing agent to reduce the incoming data in a meaningful way. For example, a data mining system could utilize these control signals to ignore certain data which is not a part of the pattern being searched for. Intelligent interfacing agents can contain the functionality of smart interfaces such as error control, but have the added feature of dynamically changing the way data passes through it.

HOLOGRAPHIC MEMORY vs. EXISTING MEMORY TECHNOLOGY

In the memory hierarchy, holographic memory lies somewhere between RAM and magnetic storage in terms of data transfer rates, storage capacity, and data access times. The theoretical limit of the number of pixels that can be stored using volume holography is $V^{2/3}/\lambda^2$ where V is the volume of the recording medium and λ is the wavelength of the reference beam. For green light, the maximum theoretical storage capacity is 0.4 Gbits/cm^2 for a page size of 1 cm x 1 cm. Also, holographic memory has an access time near 2.4 μs, a recording rate of

31 kB/s, and a readout rate of 10 GB/s. Modern magnetic disks have data transfer rates in the neighborhood of 5 to 20 MB/s. Typical DRAM today has an access time close to 10 – 40 ns, and a recording rate of 10 GB/s

Table 1: The table on the next page shows the comparison of access time, data transfer rates (readout), and storage capacity (storage density) for three types of memory; holographic, RAM, and magnetic disk

Storage Medium	Access Time	Data Transfer Rate	Storage Capacity
Holographic Memory	2.4 µs	10 GB/s	400 Mbits/cm^2
Main Memory (RAM)	10 – 40 ns	5 MB/s	4.0 Mbits/cm^2
Magnetic Disk	8.3 ms	5 – 20 MB/s	100 Mbits/cm2

Holographic memory has an access time somewhere between main memory and magnetic disk, a data transfer rate that is an order of magnitude better than both main memory and magnetic disk, and a storage capacity that is higher than both main memory and magnetic disk. Certainly if the issues of hologram decay and interference are resolved, then holographic memory could become a part of the memory hierarchy, or take the place of magnetic disk much as magnetic disk has displaced magnetic tape for most applications.

POSSIBLE APPLICATIONS

There are many possible applications of holographic memory. Holographic memory systems can potentially provide the high-speed transfers and large volumes of future computer systems. One possible application is data mining. Data mining is the process of finding patterns in large amounts of data. Data mining is used greatly in large databases which hold possible patterns which can’t be distinguished by human eyes due to the vast amount of data. Some current computer systems implement data mining, but the mass amount of storage required is pushing the limits of current data storage systems. The many advances in access times and data storage capacity that holographic memory provides could exceed conventional storage and speed up data mining considerably. This would result in more located patterns in a shorter amount of time.

Another possible application of holographic memory is in petaflop computing. A petaflop is a thousand trillion floating-point operations per second. The fast access in extremely large amounts of data provided by holographic memory systems could be utilized in a petaflop architecture. Clearly advances are needed in more than memory systems, but the theoretical schematics do exist for such a machine. Optical storage such as holographic memory provides a viable solution to the extreme amount of data which is required for petaflop computing.

ADVANTAGES

The three features of holographic memory that make it an attractive candidate to replace magnetic storage devices are redundancy of stored data, parallelism, and multiplexing. Stored data is redundant because of the nature of the interference pattern between the reference and signal beams that is imprinted into the holographic medium. Since the interference pattern is a plane wave front, the stored pattern is propagated throughout the entire volume of the holographic medium, repeating at intervals. The data can be corrupted to a certain level before information is lost so this is a very safe method of data storage. Also, the effect of lost data is to lower the signal to noise ratio so that the amount of data that can be safely lost is dependent on the desired signal to noise ratio. Stored holograms are massively parallel because the data is recorded as an optical wave front that is retrieved as a single page in one access. Since light is used to retrieve data and there are no moving parts in the detector array, data access time is on the order of 10 ms and data transfer rate approaches 1.0 GB/sec. Multiplexing allows many different patterns to be stored in the same crystal volume simply by changing the angle at which the reference beam records the hologram.

CONCLUSION

The future of holographic memory is very promising. The page access of data that holographic memory creates will provide a window into next generation computing by adding another dimension to stored data. Finding holograms in personal computers might be a bit longer off, however. The large cost of high-tech optical equipment would make small-scale systems implemented with holographic memory impractical. Holographic memory will most likely be used in next generation super computers where cost is not as much of an issue. Current magnetic storage devices remain far more cost effective than any other medium on the market. As computer systems evolve, it is not unreasonable to believe that magnetic storage will continue to do so. As mentioned earlier, however, these improvements are not made on the conceptual level. The current storage in a personal computer operates on the same principles used in the first magnetic data storage devices. The parallel nature of holographic memory has many potential gains on serial storage methods. However, many advances in optical technology and photosensitive materials need to be made before we find holograms in computer systems.