SPIE Conference on *Wave Optics and Photonic Devices for Optical Information Processing II*
August 7, 2003 Paper **5181–10**

Holography for information storage and processing

Geoffrey W. Burr

IBM Almaden Research Center,
650 Harry Road, San Jose, California 95120

ABSTRACT

We review recent progress made towards two types of holographic data storage systems. The first offers the potential for simultaneous search of an entire database by performing multiple optical correlations between stored data pages and a search argument [1,2]. This content–addressable retrieval produces one analog correlation score for each stored volume hologram. We review work we have performed on fuzzy encoding techniques, experimental demonstrations of hardware–level database searching, on the measurement of true inner–products, on architectures in which massively–parallel searches could be implemented [3], and on quantifying the inherent speed–fidelity tradeoffs [4]. The second system offers read–write, fast–access data storage. We review systems architectures for extending this high density to high capacity using phase–conjugate readout [5] and signal processing to relieve alignment and distortion constraints [6].

Keywords: volume holographic data storage, read-write holographic data storage, phase–conjugate readout, non-linear signal processing, content–addressable data storage, optical correlation, parallel search

1. INTRODUCTION TO HOLOGRAPHIC DATA STORAGE

Holographic data storage offers the benefits of large storage densities, fast parallel access, and rapid searches in large databases [7–12]. Figure 1 shows the three modes of operation in a holographic data–storage system: storage, address–based retrieval, and content–addressable searching. For storage, an entire page of information is stored at once within a photosensitive optical material. Two coherent laser beams intersect within the storage material and form an interference pattern which causes chemical and physical changes in the photosensitive medium. The information to be stored is modulated onto the object beam by passing it through a pixelated spatial light modulator (SLM). A second beam, called the reference beam, is often a simple collimated beam with a planar wavefront.

Once the hologram is recorded, either of the two beams can be used to reconstruct a copy of the other, by diffracting a small portion of the input power off the stored interference pattern. For example, in address–based retrieval, the object beam can be reconstructed by illuminating the hologram with the original reference beam. Lenses image the pixelated data page onto a matched array of detector pixels, where the bright and dark pixels can

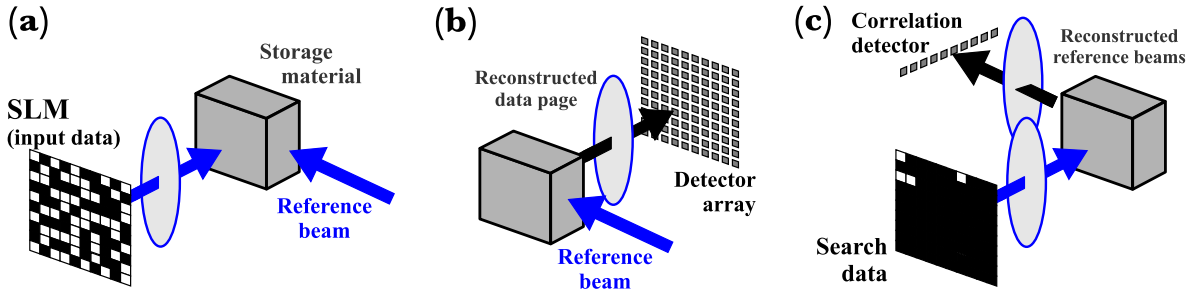

Figure 1. Holographic data storage system. **(a)** Two coherent beams, one carrying a spatial page of information, interfere within a photosensitive material to record a hologram. **(b)** Illuminating the hologram with the reference beam reconstructs a weak copy of the original information–bearing beam for capture with a detector array. **(c)** Illuminating the hologram with new page of information reconstructs all the reference beams, computing in parallel the correlation between the search data and each of the stored pages.

[1] For further information, contact G. W. Burr at *burr@almaden.ibm.com*.

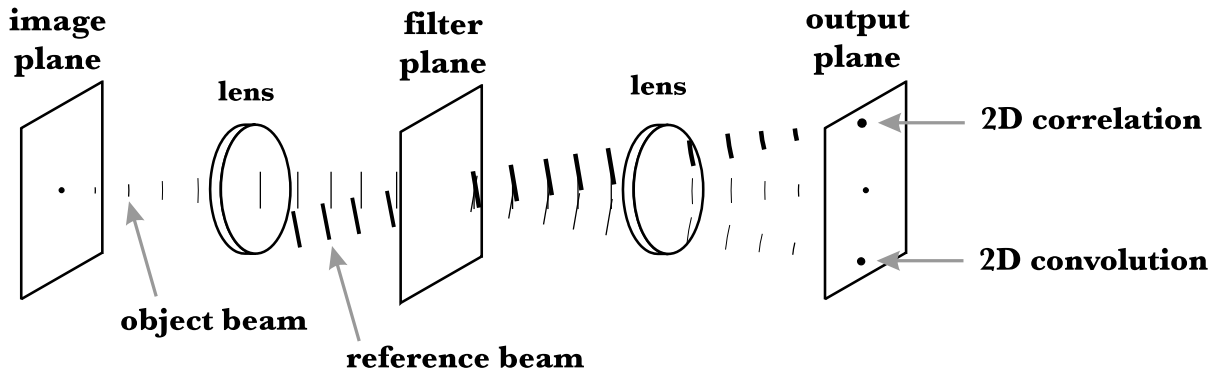

Figure 2. 4–F system, composed of two identical lenses separated by the sum of their focal lengths. Because each lens performs a spatial Fourier transform, the system performs both the 2–D correlation and the 2–D convolution between the pair of 2–D optically–input functions.

be converted back into binary data. When the hologram is stored throughout a thick storage material, then Bragg diffraction causes the strength of the reconstruction to be sensitive to changes in the angle of the reference beam. By changing this angle, multiple data pages can be stored and independently addressed (angle multiplexing). As many as 10000 holograms have been superimposed in the same 1cm^3 volume this way [13]. In addition to this high storage density, holographic data storage can also provide fast parallel readout. Since each data page can contain as many as 1 million pixels, a readout rate of 1000 pages/s leads to an output data stream of 1 Gbit/s [14].

2. VOLUME HOLOGRAPHIC CORRELATORS

In a content–addressable search the storage material is illuminated with the object beam. Consequently, all the angle–multiplexed reference beams that were used to record pages into the volume are simultaneously reconstructed by the set of stored volume holograms [1, 2, 12]. The amount of power diffracted into any one output beam is proportional to the correlation between the input data page and the associated stored data page. This correlation arises through the combination of holography and the spatial Fourier transform (FT) properties of lenses [15]. The spatial pattern that appears at the back focal plane of any lens is mathematically the 2–D Fourier transform of the spatial pattern presented in its front focal plane. This property is used by holographic data storage systems arranged in the 4F–configuration shown in Figure 2. Here an SLM is placed in the front focal plane of a lens (known as the image plane). The FT of the pattern displayed on the SLM appears in the back focal plane (filter plane). A second lens, positioned one focal length behind the storage material, performs a second Fourier transform, causing an inverted image of the original SLM data to appear at the output plane.

Multiplying the 2–D FTs of two different functions at the filter plane and then performing a second FT implements a 2–D convolution between them. In the Vanderlugt correlator [16], a hologram placed at the filter plane stores the FT of the first data page so that the second page can be presented at some later time—during content-addressable search—at the same input plane. Since it is the intensity of the interference pattern that is recorded in the storage material, the readout process essentially multiplies the FT of the displayed search page against both the FT of the stored page and the complex conjugate of the FT of the stored page. Consequently, both the 2–D convolution and the 2–D correlation are present after the FT of the second lens, as shown in Figure 2 [15]. The desired 2–D correlation function can be obtained by appropriate filtering.

With a thin hologram, only a small number of data pages can be stored, but the entire 2–D cross–correlation function can be obtained [15,17]. This method is widely used for the correlation of analog pages, such as in biometrics and target recognition [17–20], where a small stored sub-image (e.g., a particular airplane) is to be identified and located within a much larger input page (e.g., a satellite image). From the large amount of analog input data only the condensed information about possible presence and location of the target is output. Such a system should be shift–invariant, so that the position of the output correlation peak(s) follows the position(s) of the desired sub-image, with a brightness independent of position. In order to be able to clearly identify a correlation peak, it is essential to have bright, sharp peaks which stand out well from the clutter caused by non-matching inputs and background noise [17]. Beyond this, however, the particular amount of optical power in the correlation peak is fairly unimportant.

With the thick holograms that are used in the angle-multiplexed digital holographic database, Bragg-mismatch reduces the 2–D correlation between stored and searching data pages to a (roughly) 1–D function that includes the 2–D inner product [2]. Since the data pages were recorded with reference beams at different angles of incidence (angular multiplexing), correlation peaks that correspond to different reference beams (and thus different stored pages) are focussed to distinct horizontal positions in the output plane. During associative search, light is deflected from each of the stored interference patterns to the corresponding correlation peak. The power in the center of each peak represents the 2–D inner product between the searching data page and the associated stored page and therefore provides an analog measure of pattern similarity between the two pages. Consequently, if each stored page represents a data record, measuring the intensity of each peak simultaneously compares the entire stored database against the search argument [1, 12].

In contrast to target recognition, where the position and sharpness of the peaks were the critical quantities, in a holographic database it is the diffracted power in the correlation peak that measures the similarity between the data page and the search argument. Here, the system performance depends on:

- collecting sufficient signal power in the integration time given to overcome detector and other noise sources [21],
- suppressing spurious signal contributions such as scatter, crosstalk from other holograms, and signal due to the OFF pixels of the SLM [21, 22],
- reducing hologram-to-hologram variations such as changes in diffraction efficiency [21, 22], and
- the accuracy with which the detected signal power measures the 2–D inner product [21]. It is crucial that the detected signal be a function of the similarity between the search and stored data pages. As we have previously shown, this is not always the case [3].

The parallelism of content-addressable searching gives holographic data storage an inherent speed advantage over a conventional serial search, especially for large databases [2]. For instance, if an unindexed conventional "retrieve from disk and compare" software-based database is limited only by sustained hard-disk readout rate (25 MB/s), a search over one million 1KB records would take ∼40s. In comparison, with off-the-shelf, video-rate SLM and CCD technology, an appropriately designed holographic system could search the same records in ∼30ms—a 1200x improvement [2]. Custom components could enable 1000 or more parallel searches per second.

3. DATA ENCODING FOR HOLOGRAPHIC SEARCH

To allow the optical correlation process to represent a database search, the spatial patterns displayed on the SLM contain structured pixel blocks, each dedicated to a particular fixed-length field of the database [1, 2]. For example, a particular two-bit data field might be encoded by four particular pixels within the SLM page. With such an encoding, exact searches through the database can be implemented: For each ON-pixel in the input pattern, signal power is added only to the correlation peaks of stored pages in which this same pixel is also in the ON state [1]. Thus, the power in the correlation peak is a measure of the similarity between these two data pages, and matching data records are identified by simply thresholding the power. This holographic search becomes more difficult when the search page contains only a small number of arguments, because then only a small number of SLM pixels in the search page are turned ON. In this case, the weak signal from the low-power correlation peaks is often overwhelmed by background light scatter or the detector's thermal noise. Consequently, it is frequently necessary to encode data patterns with larger blocks of pixels, for example in blocks of 10×10 pixels instead of single pixels [1, 2, 21].

Errors can also arise when near matches in pixel block patterns do not correspond to near matches in encoded data value. This can cause completely unrelated records to be identified as matches when the thresholding does not work perfectly. For example, in the case of binary data encoding, a single bit error may change the data value 4 (100) into the disparate values 0 (000), 5 (101) or 6 (110). This kind of error becomes increasingly probable when many data fields are being searched for, because then the range between a full match and a total fail is being divided into a larger number of sub-levels.

3.1. Fuzzy data encoding for finding similar records

Previously, we have developed a novel data-encoding method which allows similarity or fuzzy [23] searching, by encoding similar data values into similar pixel block patterns [2]. Data values are encoded by the position of a small block of ON pixels within a column of OFF-pixels, creating a 'slider' similar to the control found on a stereo's graphic equalizer. For example, the data value 128 might be encoded as a small block of pixels, centered within a column of 256 pixels. During the search for data values near 128, the partial overlap between the input slider block and the stored slider block causes the resulting correlation peak to indicate the similarity between the input query and the stored data. This encoding scheme conserves the similarity between data values in the stored and searching page during readout, because the correlation function of two identical rectangle functions is a triangle function. Since the slope of the triangle function is a linear function, we have a linear measure of similarity between stored and searching data—as long as the linearity is preserved during correlation and detection. Since the signals are detected in intensity but add in amplitude, this linearity applies after one takes the square root of the detector signal first. This quadratic dependence can be lost when the holographic storage material is not placed exactly at the back Fourier plane of the imaging lens in the object beam [3, 21], but can be regained by using one or more stationary random phase masks in the object beam [3].

4. POTENTIAL IMPROVEMENT IN SEARCH SPEED

Because of the analog nature of the optical correlation process, low-overhead digital error correction is not available for the parallel search operation. One could use several holograms to store each database record and average the redundant correlation scores, but this incurs a large loss in storage capacity. Thus it is far more attractive to use the analog holographic database engine as a fast front-end filter to a sequential digital search engine [2]. This conserves the advantages of the holographic storage system—speed and capacity—while making it possible to satisfy the constraint of low error. For example, the system might be asked to find the 10 best matches in a database with 100,000 records. If the holographic system delivers its best 100 matches to the conventional search engine, then as long as all of the 10 best matches are in this set, the digital search engine will find them and no error will be introduced [2, 12]. This protects the holographic system from having to rank the top 10 accurately, while reducing the number of records that have to be checked by the conventional search engine by a factor of 1,000.

Because the holographic system is searching in parallel while the conventional search must process the entire database in serial, this improvement in search speed scales with the size of the database. To compare the holographic system against conventional alternatives, we plot in Figure 3 the time required to search the entire database as a function of the database size r. Each database record is assumed to contain 1000 records of 1 byte each, with a complex search (one that cannot easily be replaced by properly indexing the database) that uses a linear function of s of the 1000 fields as a search metric. To compute the search time when a microprocessor works on data pulled off a hard drive, we assume that the limiting step is the burst speed of the disk drive and that the search time follows

$$t_{\text{hard drive}} \;=\; r\,\left(\frac{1000\,\text{bytes}}{50\,\text{MB/sec}}\right). \tag{1}$$

For the DRAM curves, we assume that the required 1 GByte of memory (45ns random search, 400MHz burst transfer rate) is feeding a 1 GHz microprocessor with a small on–chip cache. For small s, it is more economical to do a random search for each of the s bytes and the search time follows

$$t_{\text{DRAM, small } s} \;=\; r\,\left(s\,(45\text{ns}) \;+\; \frac{4\,s\;+\;\log_2 r}{1000\,\text{MHz}}\right). \tag{2}$$

As s becomes large, it becomes quicker to read the entire 1kByte record into the local cache and then access the needed bytes, so the time becomes

$$t_{\text{DRAM, large } s} \;=\; r\,\left((45\text{ns}) \;+\; \frac{1\,\text{kB}}{400\,\text{MHz}} \;+\; \frac{5\,s\;+\;\log_2 r}{1000\,\text{MHz}}\right). \tag{3}$$

(Note that this somewhat unreasonably implies an 8–bit–wide data bus.) In contrast, the search time required by the holographic database engine is the optical exposure time, followed by the same DRAM search performed on a smaller number of records, r'.

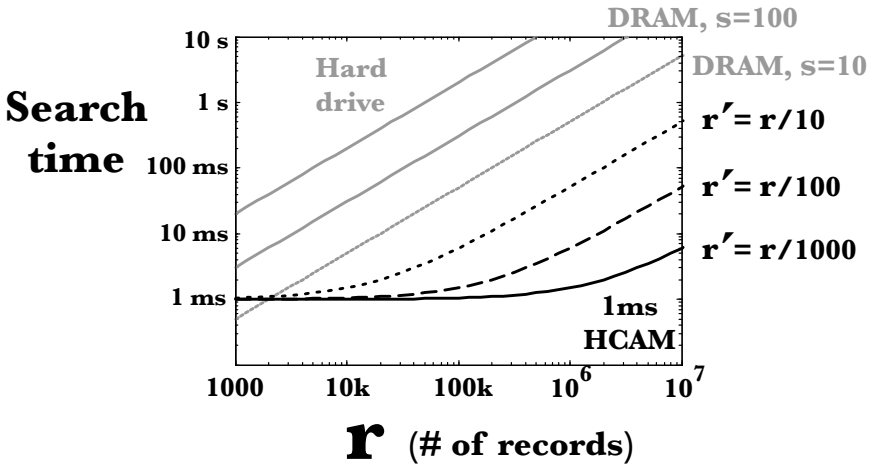

Figure 3. Plot of time required to search an entire database versus the number of database records, r. Each record is assumed to contain 1000 one–byte fields, of which s fields are to be extracted, weighted, and summed to compute the search metric. Gray lines for conventional technologies (Hard drive, DRAM with s=100 and s=10) are compared to the potential of a holographic search engine (black lines). The impact of r', the number of records passed by such a search engine to a successive DRAM search is shown.

5. RELATING MEASURED CORRELATION SCORES TO DATABASE RECORD SIMILARITY

With the combination of the holographic search engine and a conventional digital search engine, *most* of the time no errors will be introduced by the holographic system. However, there exists a small statistical chance that one of the matching records will fail to be included within the set of records passed to the electronic system. One can reduce this probability of error by increasing the number of matches passed to the electronic system; however, this forces the conventional serial search to work on more records and reduces the speed advantage provided by the optical system. Previously, we have described how to compute this probability of error and consider the resulting tradeoffs between the desirable performance metrics of the holographic content–addressable system, including probability of error, search speed, the number of holograms in the database, and the number of fields per holographic record [4].

The only output of the holographic search engine is the set of r correlation scores, one for each stored hologram. (This is the optical power in the correlation peak measured at the position at which each hologram's reference beam came to focus during recording: the measured 2–D inner product between the search page and that particular stored data page.) For simplicity, we assume that there is a one–to–one correspondence between holograms and database records.

To calculate the probability with which the analog optical system makes errors, we first relate the similarity between each database record and the search query to the optical correlation scores that would be expected in the absence of noise. We can then see how noise perturbs these correlation scores so that matching records may possibly be overlooked. In Figure 4, we graphically represent these relationships through a series of plots. The bottom plot, Figure 4(a), details the expected scores; the central plot, Figure 4(b), relates these expected scores to the actual measured scores; and the righthand plot, Figure 4(c), shows the statistical variation of measured scores due to the random noise. Parts (a) and (b) share a common horizontal axis; Parts (b) and (c) share a common vertical axis. Part (a) represents the number of occurrences of each expected score, given the particular database and query under consideration. This can be thought of as the result of reducing the multi–dimensional database to a single dimension by applying the query. The horizontal axis, common to both parts (a) and (b), represents the relative similarity between the database records and the search query. For the holographic database engine, this relative similarity can be simply the SLM area in common between the search query page and the stored database page.

In the absence of noise, all records with a particular expected score or relative similarity would produce identical measured correlation scores. Because the readout light is coherent and the holograms are weak (the first Born approximation applies) [3, 24], the relationship between the expected score in SLM pixels and the measured score in

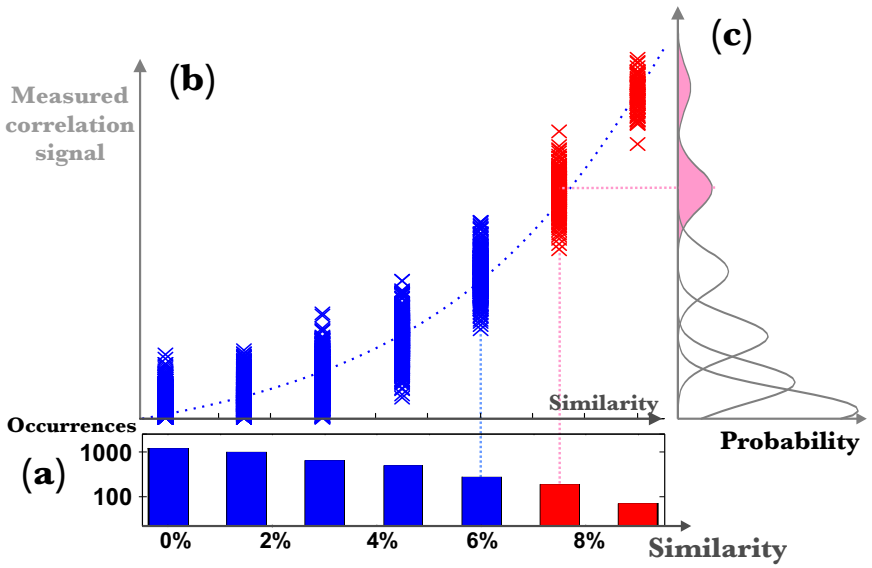

Figure 4. Relationship between expected score (the similarity between database records and the search query in terms of SLM area in common) and the actual score measured by the holographic search engine in the presence of noise (in detected photons). Part(a) describes the database and query in terms of the distribution of expected scores; Part(b) shows the mapping between expected and measured scores; and Part(c) shows the distribution of measured scores as dictated by the probability density function of the random noise.

photons is quadratic. In the presence of noise, however, the measured score may be higher or lower than this average score, varying as the probability density function (PDF) of the overall noise. Note that at each horizontal position, there are as many X's as there are database records with this expected score (this can be read off the vertical axis of part (a)). On the right, in part (c), we show the PDFs for the measured scores. The vertical axis, common to both parts (b) and (c), represents measured correlation score in units of detected photons. In any experimental apparatus, these measured correlation scores are the only available information from which to determine which records matched the search query. A threshold must be applied at some value of measured score in order to pick out the matching records (e.g., those with high scores) from the rest.

6. TRADEOFF BETWEEN SNR AND THE NUMBER OF NEAR–MATCHES

As described above, if the holographic search engine is feeding its results into a conventional electronic system, then an error only occurs if one of the matching records fails to be selected by the cutoff threshold applied to the measured correlation score. The probability of making such an error can be greatly reduced by simply reducing the threshold, at the cost of increasing the number of records passed. These near–matches increase the workload of the electronic processor and thus reduce the speed advantage of the holographic search engine.

When a high probability of error is acceptable then only the matching records need to be passed to the electronics, but when low probability of error is required then many superfluous but near–matching records must be passed.

The number of superfluous near–matches passed depends not only on SNR and desired P_e, but also on the particular database search being attempted. Only those records with expected scores near the transition(s) between matching and non–matching records have a large impact on the probability of error, since these are the records most likely to be mis–classified. When the SNR gets too low, the holographic search engine is essentially useless, passing on almost all of the records. As we have previously shown, the probability of a search error, P_e, and the fraction of database passed to the electronics can be connected to the other performance parameters of the holographic system such as the number of holograms, the number of database fields per holographic data page, and the optical search speed [4].

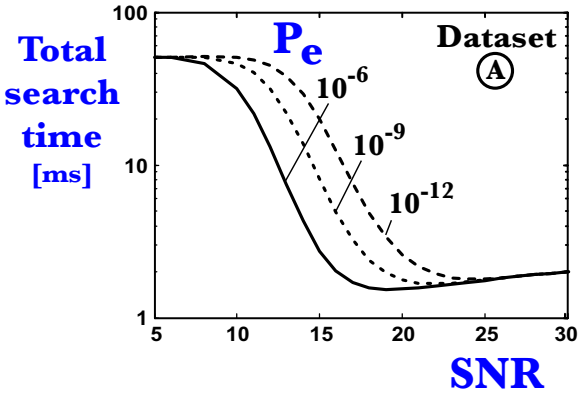

Figure 5. Plot of search time vs. SNR, showing the impact of lower SNR (many records must be passed to the electronic processor) and higher SNR (the optical exposure is longer than it needs to be). The ratio between the minimum and maximum search times represents the performance improvement offered by the holographic database engine.

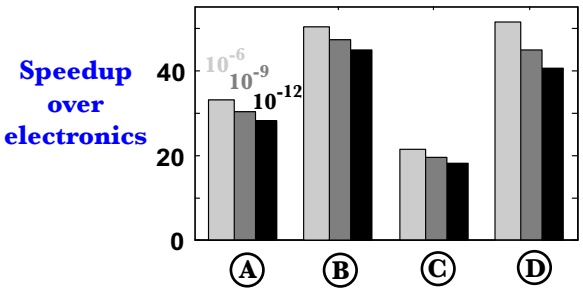

Figure 6. Factors by which the holographic database engine outperforms a conventional search, for the four database examples listed in Figure 7 and three different probability of error targets.

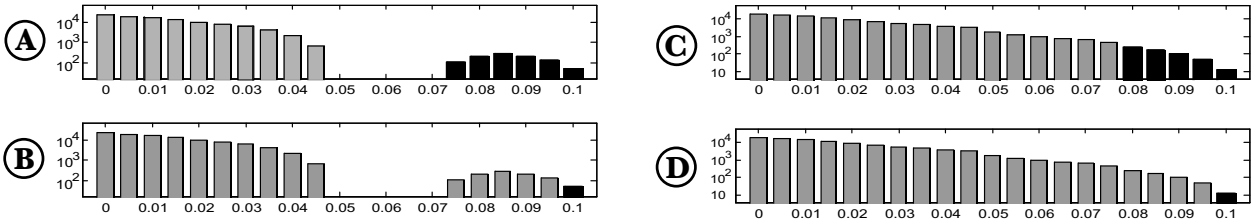

Figure 7. The four example database searches used, showing occurrences on a log scale as a function of the expected similarity. Black bars represent matching expected scores, while gray bars represent the range of scores for non–matching records. The total number of records r is 100,000 for each. Note that even redefining the subset of matching records can have a significant impact on the error–vs.–overhead tradeoff.

7. REALIZABLE IMPROVEMENT IN SEARCH SPEED

There is a variation in the probability of error from search to search, depending on the boundary between the matching and near–matching records [4]. Thus one can achieve a larger performance improvement for easier searches, because one can use a shorter optical exposure time or pass a smaller number of records to the electronic system. Interestingly, however, one cannot do both of these things: shorter exposure time implies less signal and thus lower signal–to–noise ratio, but lower signal–to–noise ratio implies that more records need to be passed in order to maintain the target probability of error.

These two trends are shown for a particular dataset in Figure 5 for three different P_e targets [4]. The ratio between the minimum and maximum search times represents the performance improvement offered by the holographic database engine, and these speedup factors are shown for four particular database searches in Figure 6. The four database searches are defined in Figure 7, which shows the distribution of expected signals and identifies which records are considered to be matching. Depending on the difficulty of the search, the performance offered by the holographic database engine for these database queries ranges from 20 to 50. Note that, as suggested in Figure 3, this factor should increase even further for larger r, until the subsequent electronic search through the smaller subset r' begins to dominate the total search time.

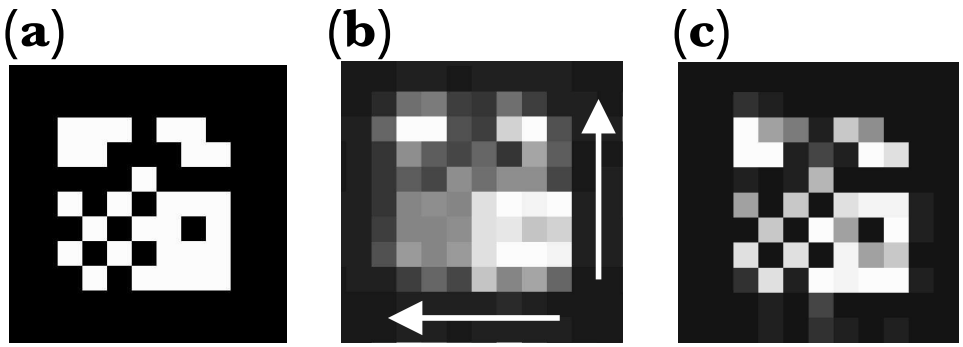

Figure 8. A 9×9 pixel pattern is imaged from SLM to CCD (a) under perfect conditions; (b) with half–a–pixel offset in both x and y; (c) after post–processing with the shift–compensation algorithm [6].

8. POST–PROCESSING TO COMPENSATE FOR PIXEL SHIFTS

The high density and fast readout offered by address–based volume holographic data storage are due in large part to the arrangement of data into large pixelated pages [11]. But to retrieve any stored data, the pixel array imposed by the input spatial light modulator (SLM) must be accurately delivered to the array of detector pixels. Displacement of individual pixel images away from their target detector pixels (because of magnification error, misalignment, optical distortion, or material shrinkage [25]) quickly leads to uncorrectable levels of error.

Linear signal processing techniques have had success deblurring conventional 1–D channels found in modern communication and data storage devices. However, these techniques are not directly applicable to holographic storage, because the optical detection in holographic storage is inherently nonlinear. With coherent light, the intermingling of signal from neighboring pixels takes place in the amplitude domain [26], yet the signals accessible to a post–processing algorithm are the spatial integrals of the square of this optical amplitude (i.e., intensity) over the area of individual pixels [27]. Because of the spatial integration, one cannot simply take the square root to return to incident amplitude. In addition to pixel blur, once data pages reach a megapel (1024×1024 pixel) in size, a significant portion of the SNR budget is consumed simply by residual optical distortion: one portion of the page must always be partly misaligned in order to bring another portion into optimal alignment [14, 28]. To jointly address these two problems, we have recently derived an algorithm that can compensate for both optical distortion and misalignment, correcting a moderate pixel blur in the presence of a significant pixel offset [6].

The algorithm comes directly from the detection physics in holographic data storage. Consider the readout signal received at a detector pixel when the incoming data page is shifted such that two SLM pixel images are contributing (the correct SLM pixel, and one neighbor). In the simplest 1–D case, we can decompose this detected signal r_2 into linear contributions from the two SLM pixel intensities p_1, p_2, and a nonlinear factor through their constructive interference, as

$$r_2 = p_2 H_{00}(\sigma) + 2\sqrt{p_1 p_2} H_{01}(\sigma) + p_1 H_{11}(\sigma). \tag{4}$$

The weights $H_{00}(\sigma), H_{11}(\sigma)$ and $H_{01}(\sigma)$ represent the normalized signal integrated by the detector pixel from the correct SLM pixel alone, the signal from the neighboring SLM pixel alone, and the additional contribution when both SLM pixels are present, respectively. This equation can be inverted to iteratively to solve for each pixel's correct value (p_2) from the just–processed neighbor pixel (p_1) and the received data signal (r_2). At each pixel, we take the measured signal, subtract the portion that belonged to the previous pixel, subtract a further portion due to interference, and then add in the missing signal that should have been here but which actually fell into the next pixel. Because the point–spread function in the DEMON II system [28] is dominated by diffraction effects, the 2–D data page can be processed by using this simple 1–D algorithm repeatedly, first on all the rows, and then on the columns [6].

In Figure 8(a), we show a small 9×9 pixel block as it should ideally be received. Figure 8(b) shows the same pattern imaged through DEMON II when the SLM is shifted a half–pixel in both x and y. The DEMON II platform pixel–matches megapel pages through an aperture of 1.36 D_N (1.7 x 1.7mm^2 aperture, f=30mm, λ=532nm, SLM δ= 12.8 microns). In Figure 8(c), we show this data after post–processing with the shift–compensation algorithm: the original pixel pattern is effectively recovered [6]. To process each block of pixels, the algorithm combines the

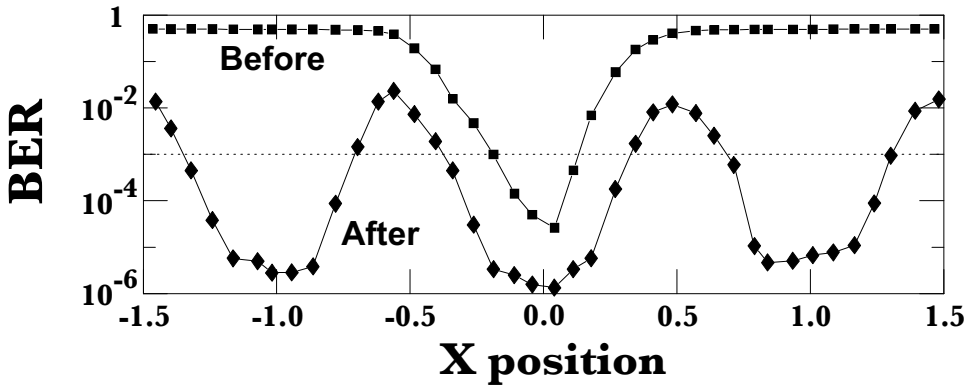

Figure 9. Dependence of raw–BER (with an 8:12 modulation code [29]) before and after shift–compensation post–processing, as a function of x shift [6].

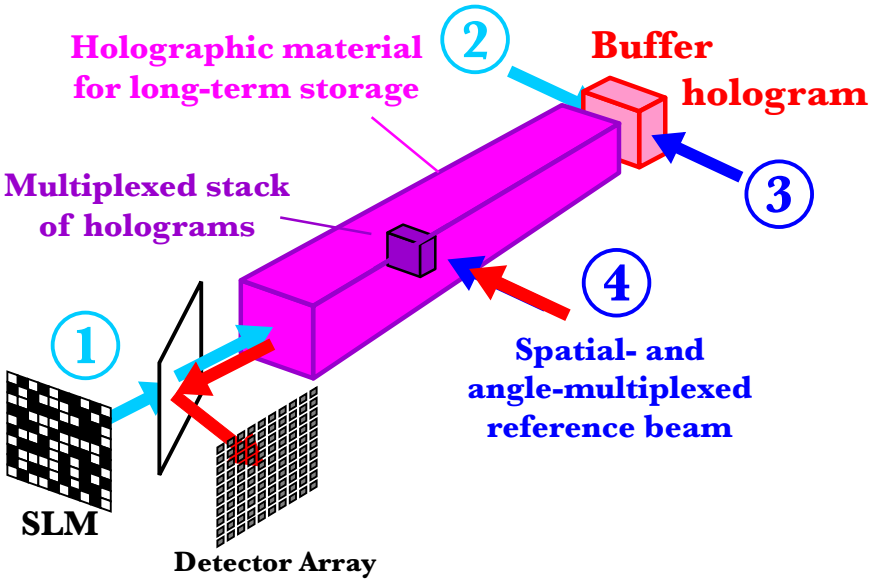

Figure 10. Phase–conjugate holographic storage system using a buffer hologram A temporary buffer hologram is recorded by an object beam containing the data from the SLM (1) and a reference beam (2). This hologram is illuminated with a phase–conjugate beam (3), reconstructing the phase–conjugate of the original object beam, which is then stored permanently with a spatial– and angle–multiplexed reference beam (4) [5].

dynamic global shifts, as measured by dedicated fiducial marks, with the static local baseline offsets taken from a lookup table [6]. Assuming that 10^{-3} is the maximum acceptable raw bit–error–rate (BER) that can be corrected by error–correction codes of moderate overhead [28], the shift–compensation algorithm increases the position tolerance of the DEMON II tester from ±16% to ±40% of the pixel pitch [6]. This is demonstrated experimentally (with images) in Figure 9.

9. PHASE–CONJUGATE READOUT

The success of the recent holographic storage experiments (such as the DEMON II platform [28]) has been due jointly to excellent imaging fidelity (pixelated data arrives at the right detectors) and tight focusing of the object beam (holograms can be stored using very little volume). These two features were made possible by optical design: minimizing the optical aberrations, particularly distortion, of the short focal length optics. To scale fast–access holographic storage to high capacity, however, this same high density must be achieved at many storage locations, and without moving the storage media. The correspondingly greater demands on optical imaging performance soon

limit the capacity achievable along this path to commercially uninteresting limits. However, several researchers have long proposed bypassing these imaging constraints with "phase–conjugate" readout.

Once the light from the spatial light modulator has been recorded using a reference beam, the resulting hologram can be reconstructed with a phase–conjugate or "time-reversed" copy of the original reference beam. The wavefront diffracted by the phase–conjugate readout beam then retraces the path of the incoming object beam in reverse, cancelling out any accumulated phase errors from lens aberrations or material imperfections. This allows data pages to be retrieved with high fidelity using an inexpensive lens [30, 31], or even without imaging lenses for an extremely compact system [32, 33].

However, two uncertainties prevented earlier work from proceeding. First, researchers were worried that imperfections in producing the phase–conjugate reference beam would introduce errors in each retrieved data page. In the simplest case, the phase–conjugate reference beam is a separate light beam that is carefully aligned to propagate in exactly the opposite direction to the original reference beam. However, even minor differences between the two beams will distort the reconstructed data pages. Alternatively, an extremely accurate phase–conjugate beam can be produced by a self–pumped phase–conjugate mirror [34]. Recently, we showed that a beam reflected from a phase–conjugate mirror could retrieve pages containing one million pixels onto a camera with very few bit errors, providing at least one potential solution to this first uncertainty [35].

The second concern was that many pairs of phase–conjugate reference beams would be needed to read the many different holograms recorded within the same volume—-and maintaining these beams over long periods of time would be impossible from a practical point of view. This problem also kept researchers from using the phase–conjugate mirror or PCM, since the barium titanate crystal takes some time to respond when the input beam changes.

9.1. Buffer hologram

To solve this problem, we have proposed [5]—and are currently completing the evaluation of—a novel architecture that allows phase–conjugation and multiplexed holographic storage to co–exist. The technique involves separating the phase–conjugation and hologram storage processes into two successive steps by using a 'buffer' hologram [5], as shown in Figure 10. Data to be recorded are modulated onto the object beam (1) with the SLM and focused into a long storage crystal. The object beam travels down the crystal, confined by total internal reflection, and passes into a buffer crystal, where it interferes with beam (2) and records a hologram. This hologram is then immediately read with beam (3), the phase–conjugate of beam (2), reconstructing a phase–conjugate object beam which travels back into the storage crystal. This new object beam can now be recorded, and then later reconstructed, with beam (4) at one of the storage locations. This permits all the same angle–, phase–code–, wavelength–, and spatial–multiplexing approaches used for conventional volume holograms [7, 8].

Holograms can then easily be multiplexed at a large number of separate storage locations using only one SLM and one detector array. The buffer hologram technique still requires a pair of phase–conjugate reference beams, although only a single pair which do not need to ever change angle. Thus the self–pumped phase–conjugate mirror can be used, providing accurate phase–conjugation and adaptation to system misalignments without requiring the storage system to repeatedly wait for the PCM reflectivity to build up. Alternatively, the single phase–conjugate reference beam needed for the buffer hologram could be generated by careful alignment of a counter–propagating beam. This might have advantages over the self–pumped PCM, for example when implementing the buffer hologram in a wavelength–multiplexed system or when high modulation depth is required while recording the buffer hologram. Other advantages of the buffer hologram system include the ability to align the detector array to the SLM through the buffer hologram without sacrificing storage capacity. Finally, the strong buffer hologram can be monitored during recording by diverting, at the beamsplitter, a small portion of the returning phase–conjugate object beam to the detector while continuing to pass most of the object beam power from the SLM to the media.

9.2. Materials requirements

The long–term storage material in a phase–conjugate/buffer hologram system must provide highly sensitive, non–volatile storage. One candidate is two–color, gated volume holography in $LiNbO_3$ (see next section). The object and reference beams use long wavelength light (say, red or IR) which the crystal only absorbs in the presence of short wavelength gating light (green). The storage crystal can then be made extremely long, the gating light used to activate storage locations, and stored holograms read out without erasure.

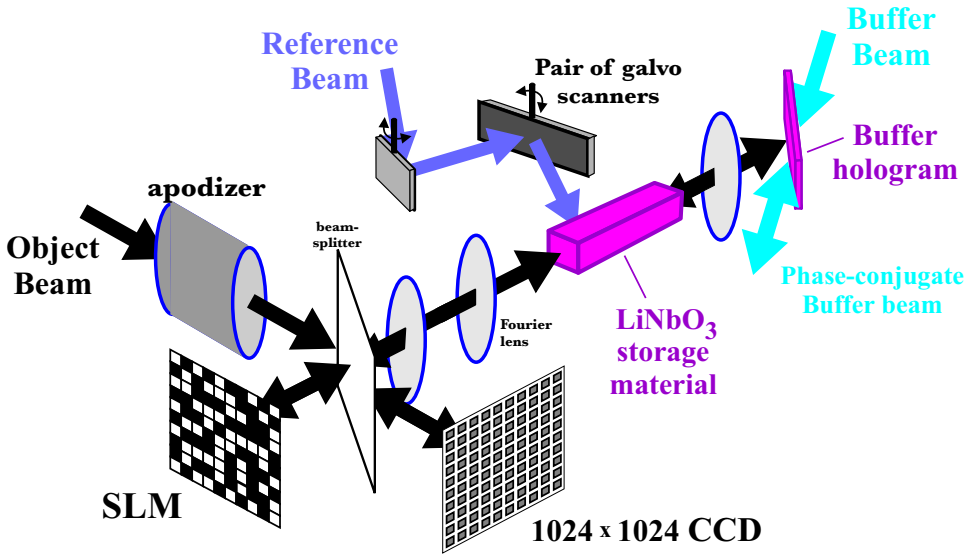

Figure 11. DEMON III holographic storage platform, for testing the use of a buffer hologram with multiplexed phase–conjugate readout.

Since this technique for multiplexed phase–conjugate holograms records two holograms for each stored data page, it would seem to inherently slow down the recording process. However, once the diffraction efficiency of the buffer hologram exceeds the power efficiency of the original object beam (typically ∼1–10%), then the recording of the storage hologram is actually accelerated. The buffer hologram does require a dynamic holographic material with high sensitivity, so that each new data page completely and rapidly overwrites the previous one. However, dynamic range for multiple holograms, dark storage lifetime, hologram thickness, and optical quality are less important, and could be traded off during material optimization for more sensitivity. Low scattering and uniform spatial frequency response are still needed, however, and the material must be able to tolerate a large number of read–write–erase cycle without degradation. There are several read–write materials whose strengths and weaknesses fit well here, including photorefractive polymers [36,37], bacteriorhodopsin [38], and fast photorefractive materials such as strontium barium niobate (SBN).

Erasure of the buffer hologram is preferably induced externally, either with electric field or an incoherent erase beam. This is particularly true when using the self–pumped PCM, because while the weak return beam is being used at the buffer material, the strong beam must also be present. If both beams are present during recording, then the modulation depth at the buffer material is greatly reduced. When both are present during readout and the material is not gated, then the erasure of the buffer hologram is greatly accelerated. If the buffer material is gated, then the PCM can be used upon readout and the strong pump beam safely ignored. For an ungated buffer material, however, adding buffer readout power—to get a stronger phase–conjugate object beam for transfer into the long–term hologram—is actually disadvantageous. The benefits of an initially stronger object beam are outweighed by the faster buffer erasure, so that increases in buffer readout power end up reducing the amount of object beam *energy* reaching the long–term hologram. The best option, in the absence of a gated buffer material, is to write a strong hologram and read it with as weak a readout beam as is feasible. It is also important to consider the impact of the erasure induced in the long–term storage material during transfer from the buffer, and the need to transfer holograms of varying strength within a recording schedule.

9.3. DEMON III test platform

We have built yet another test platform, called DEMON III, to evaluate the phase–conjugate/buffer hologram technique, using conventional, one–color lithium niobate. The optical layout is shown in Figure 11. A beam from a Coherent DPSS 532nm laser is expanded and split with various beam expansion optics to create three beams: an object beam for illuminating the SLM, a reference beam for conventional spatial– and angle–multiplexing, and a third beam for writing the buffer hologram. The SLM and CCD are identical to those in the DEMON II platform. In DEMON III, however, a 1:1 relay lens delivers the SLM image to the input plane of a short–focal length Fourier

transform lens, which then focuses the object beam into the $2\times2\text{mm}^2$ aperture of a long $LiNbO_3$:Fe bar (cut for the 90° geometry). The object beam expands out of the back end of the $LiNbO_3$ onto the buffer hologram material. The phase–conjugate reference beam for the buffer hologram can either be provided by a self–pumped phase–conjugate mirror in $BaTiO_3$ [5, 34], or by a second beam aligned to be counter–propagating. After the object beam is phase–conjugated, it can be stored in the long $LiNbO_3$ bar with the reference beam. A pair of galvo mirrors provides control over beam position along the bar as well as horizontal incidence angle.

To date, we have demonstrated pixel–matching of megapel pages in this platform through a simple biconvex lens of low F/#, despite significant aberrations in the image before phase–conjugation. The buffer hologram is written then immediately read by a pair of beams carefully aligned to be phase–conjugate, simply by maximizing the SNR of the resulting holograms with a random phase mask in the object beam. These experiments, performed without the lithium niobate crystal in place, have used photorefractive polymer material [39], bacteriorhodopsin [40], and SBN as the buffer hologram material. The remaining experimental difficulty is the low diffraction efficiencies attained with all three of these materials, approximately 0.1% at best, which is not quite sufficient for transfer of these phase–conjugate object beams into the long–term $LiNbO_3$:Fe crystals.

9.4. Extended shift–compensation algorithm

The shift–compensation procedure described above can relax the tight constraints on page registration, optical distortion, and material shrinkage that currently hamper page–oriented holographic storage systems. Since a shift of an integral number of pixels requires only careful bookkeeping, improvement of this algorithm to tolerance of a ±50% pixel shift would imply that the data could be retrieved as long as it falls somewhere on the detector array. This would provide many exciting opportunities to push system design (disk rotation rate), to relax mechanical constraints (disk wobble or readout pulse timing tolerances), and to tolerate material imperfections (shrinkage and expansion, whether induced thermally or optically).

Recently we have modified the shift–compensation algorithm so that it can successfully correct for arbitrary page misalignments. This modification requires that the SLM and CCD pattern have different magnifications, which we obtain for DEMON III platform by the simple expedient of removing the magnification optics. The SLM pattern, with 12.8μm pixel pitch, is then reconstructed by the phase–conjugate readout onto the CCD camera of 12μm pitch (6.25% magnification error). Lines that are 15 pixels apart on the SLM can be perfectly captured by CCD pixel columns that are 16 pixels apart. Everything in–between must be interpolated by the shift–compensation algorithm. The bookkeeping is somewhat intricate, since the signal received at pixel 200,200 is now no longer associated with the SLM pattern that was transmitted at pixel 200,200, and all the data must be correctly moved to the appropriate pixel location in order to allow decoding and BER analysis.

Figure 12 shows a portion of a 1×1 pixel chessboard pattern, first as it is received by the CCD camera (part (a)), and then after correction with the modified shift–compensation algorithm. In Figure 13 shows BER as a function of shift for pages imaged on DEMON II without shift compensation, pages imaged on DEMON II with the original shift compensation algorithm, and holograms reconstructed with the modified shift compensation algorithm on DEMON III with the 6.25% magnification error described above. The presence of the magnification error means that all possible local shifts are represented on each data page, independent of global shift. Thus the modified algorithm provides true invariance to page misalignment, magnification, and optical distortion. Although isolated pages with BERs as low as 2×10^{-5} have been retrieved in the presence of this magnification error, we continue to work to improve the robustness and decrease the BER that this modified shift–compensation algorithm can produce.

We anticipate that the successful use of phase–conjugation and shift–compensation in holographic storage will enable compact and affordable high–capacity systems, with only a moderate increase in the overall system complexity. However, such systems will require a recording material that supports both read–write access and non–volatile storage.

10. CONCLUSIONS AND OUTLOOK

Holographic data storage has several characteristics that are unlike those of any other existing storage technologies. Most exciting, of course, is the potential for data densities and data transfer rates substantially exceeding those of magnetic data storage. In addition, as in all other optical data storage methods, the density increases rapidly with decreasing laser wavelength. In contrast to surface storage techniques such as CD–ROM, where the density is inversely proportional to the square of the wavelength, holography is a volumetric technique, making its density proportional to one over the third power of the wavelength. In principle, laser beams can be moved with no mechanical

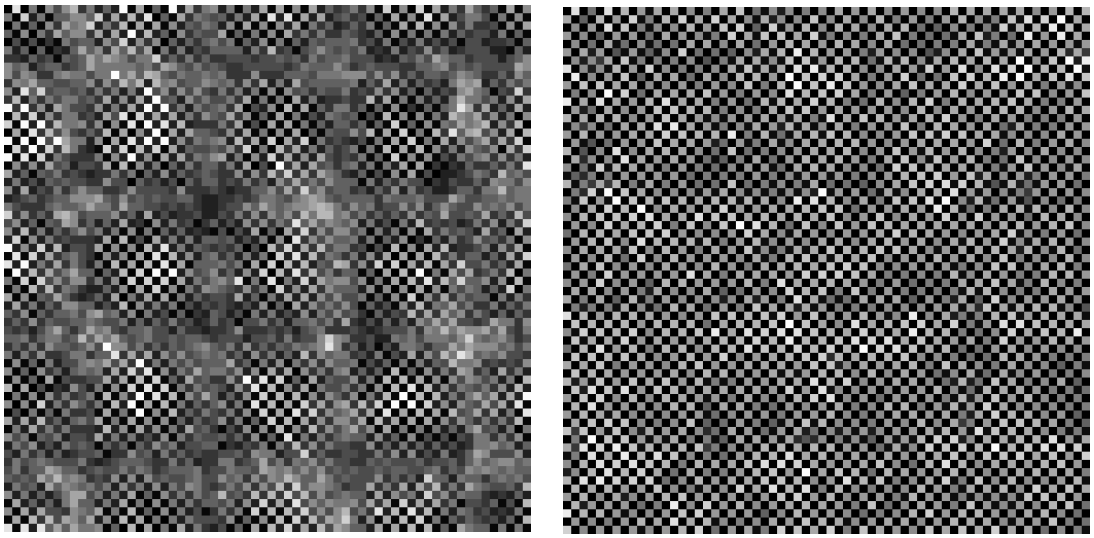

Figure 12. 1×1 pixel chessboard pattern (a) after phase–conjugate reconstruction onto CCD detector with 6.25% magnification error (slips half a pixel every 8 pixels), and (b) after post–processing with modified shift–compensation algorithm.

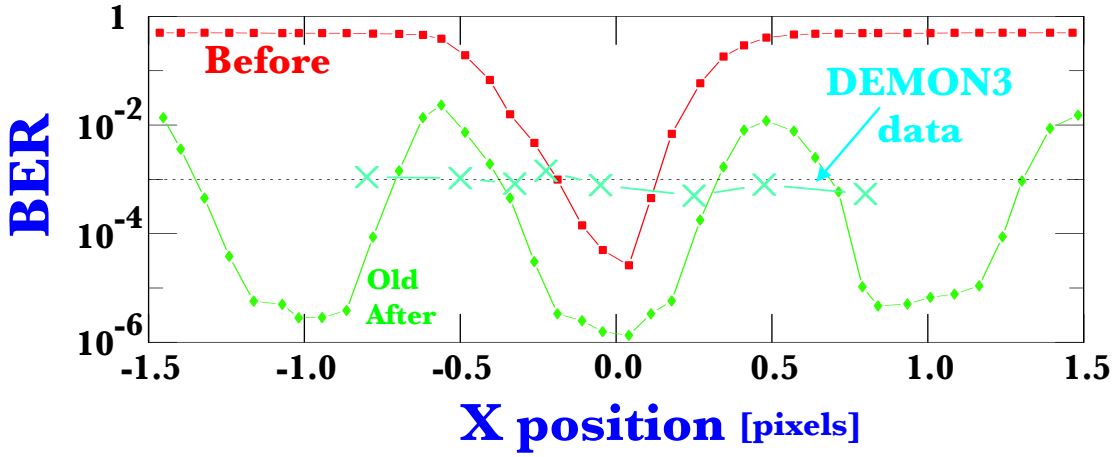

Figure 13. BER (with 8:12 modulation code) as a function of page misalignment without shift–compensation, with the original shift–compensation algorithm, and with the modified shift–compensation algorithm. With the original algorithm, the poor performance at a half–pixel shift means that the page-wide BER becomes unacceptable at these shifts. With the magnification error that the modified algorithm corrects, some portions of the page are at half–pixel shifts while others are aligned directly on pixels and the BER does not change with global page shift.

components, allowing access times of the order of 10 microseconds, faster than any conventional disk drive will ever be able to randomly access data.

With its inherent parallelism, volume-holographic content addressable data storage is an attractive technique for searching large databases with complex queries. Because a holographic database engine searches an entire database with a single optical exposure, it can potentially search massive databases orders of magnitude faster than conventional alternatives. However, the analog nature of the measured optical correlations implies that records that should be selected may be inadvertently overlooked due to random noise and deterministic variations. The probability that such a parallel search is erroneous can be managed by using a conventional electronic search engine in combination with the optical search engine. By carefully optimizing the number of records passed and the signal–to–noise ratio with which the optical search is performed, improvements in search speed of 20 to 50 can be achieved

for databases of 100,000 records while keeping the probability of missing even a single record at 10^{-12}.

The research efforts of the last few years have demonstrated that holographic storage systems with desirable properties can be engineered and built in the laboratory. However, existing and other developing storage technologies also continue to evolve and improve at a tremendous pace, making the next few years crucial for holographic storage. The next steps are to optimize the storage media, to demonstrate these systems outside the laboratory environment, and to design and build systems that are cost–competitive with existing technologies.

REFERENCES

1. B. J. Goertzen and P. A. Mitkas, "Volume holographic storage for large relational databases," *Optical Engineering* **35**(7), pp. 1847–1853, 1995.
2. G. W. Burr, S. Kobras, H. Hanssen, and H. Coufal, "Content–addressable data storage by use of volume holograms," *Applied Optics* **38**(32), pp. 6779–6784, 1999.
3. F. Grawert, G. W. Burr, S. Kobras, H. Hanssen, M. Riedel, C. M. Jefferson, M. Jurich, and H. Coufal, "Content–addressable holographic databases," in *Critical technologies for the future of computing*, vol. 4109 of *Proceedings of the SPIE*, pp. 177–188, Jul 2000.
4. G. W. Burr, G. Maltezos, F. Grawert, S. Kobras, H. Hanssen, and H. Coufal, "Using volume holograms to search digital databases," in *Conference on Three– and Four–Dimensional Optical Data Storage*, vol. 4459 of *Proceedings of the SPIE*, pp. 311–322, Jul 2001.
5. G. W. Burr and I. Leyva, "Multiplexed phase-conjugate holographic data storage using a buffer hologram," *Optics Letters* **25**(7), pp. 499–501, 2000.
6. G. W. Burr and T. Weiss, "Compensation of pixel misregistration in volume holographic data storage," *Optics Letters* **26**(8), pp. 542–544, 2001.
7. D. Psaltis and F. Mok, "Holographic memories," *Scientific American* **273**(5), p. 70, 1995.
8. J. H. Hong, I. McMichael, T. Y. Chang, W. Christian, and E. G. Paek, "Volume holographic memory systems: techniques and architectures," *Optical Engineering* **34**, pp. 2193–2203, 1995.
9. J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science* **265**, p. 749, 1994.
10. D. Psaltis and G. W. Burr, "Holographic data storage," *IEEE Computer* **31**(2), pp. 52–60, 1998.
11. J. Ashley, M.-P. Bernal, G. W. Burr, H. Coufal, H. Guenther, J. A. Hoffnagle, C. M. Jefferson, B. Marcus, R. M. Macfarlane, R. M. Shelby, and G. T. Sincerbox, "Holographic data storage," *IBM J. Research and Development* **44**, pp. 341–368, May 2000.
12. G. W. Burr and P. Mitkas, *Holographic Data Storage*, ch. Volume holographic correlators. Springer–Verlag, Berlin, 2000. ed., H. Coufal and D. Psaltis and G. Sincerbox.
13. G. W. Burr, *Volume holographic storage using the 90° geometry.* PhD thesis, California Institute of Technology, Pasadena, Calif., 1996.
14. R. M. Shelby, J. A. Hoffnagle, G. W. Burr, C. M. Jefferson, M.-P. Bernal, H. Coufal, R. K. Grygier, H. G. . Günther, R. M. Macfarlane, and G. T. Sincerbox, "Pixel–matched holographic data storage with megabit pages," *Optics Letters* **22**(19), pp. 1509–1511, 1997.
15. J. W. Goodman, *Introduction to Fourier Optics*, McGraw–Hill, 1968.
16. A. Vander Lugt, "Signal detection by complex spatial filtering," *IEEE Transactions on Information Theory* **IT–10**, pp. 139–145, 1964.
17. S. H. Lee, ed., *Optical Information Processing–Fundamentals*, Springer–Verlag, Berlin, 1981.
18. E. G. Paek and D. Psaltis, "Optical associative memory using Fourier–transform holograms," *Optical Engineering* **26**(5), pp. 428–433, 1987.
19. H.-Y. S. Li, Y. Qiao, and D. Psaltis, "Optical network for real time face recognition," *Applied Optics* **32**(26), pp. 5026–5035, 1993.
20. A. Pu, R. Denkewalter, and D. Psaltis, "Real time vehicle navigation using a holographic memory," *Optical Engineering* **36**(10), pp. 2737–2746, 1997.
21. S. Kobras, "Associative recall of digital data in volume holographic storage systems," Master's thesis, Technische Universitat Munchen, May 1998.
22. G. A. Betzos, A. Laisne, and P. A. Mitkas, "Improved associative recall of binary data in volume holographic memories," *Optics Communication* **171**(1–3), pp. 37–44, 1999.
23. L. A. Zadeh, "Fuzzy sets," *Information control* **8**, pp. 338–353, 1965.
24. H. J. Coufal, D. Psaltis, and G. Sincerbox, eds., *Holographic Data Storage*, Springer–Verlag, 2000.
25. R. M. Shelby, D. A. Waldman, and R. T. Ingwall, "Distortions in pixel-matched holographic data storage due to lateral dimensional change of photopolymer storage media," *Optics Letters* **25**(10), pp. 713–715, 2000.
26. V. Vadde and B. V. K. Vijaya Kumar, "Channel modeling and estimation for intra–page equalization in pixel–matched holographic data storage," *Applied Optics* **38**(20), pp. 4374–4386, 1999.

27. M.-P. Bernal, G. W. Burr, H. Coufal, and M. Quintanilla, "Balancing inter–pixel crosstalk and thermal noise to optimize areal density in holographic storage systems," *Applied Optics* **37**(23), pp. 5377—5385, 1998.
28. G. W. Burr, C. M. Jefferson, H. Coufal, M. Jurich, J. A. Hoffnagle, R. M. Macfarlane, and R. M. Shelby, "Volume holographic data storage at an areal density of 250 Gigapixels/in^2," *Optics Letters* **26**(7), pp. 444–446, 2001.
29. G. W. Burr, J. Ashley, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, and B. Marcus, "Modulation coding for pixel–matched holographic data storage," *Optics Letters* **22**(9), pp. 639–641, 1997.
30. F. Ito, K.-I. Kitayama, and H. Oguri, "Compensation of fiber holographic image distortion caused by intrasignal photorefractive coupling by using a phase–conjugate mirror," *Optics Letters* **17**(3), pp. 215–217, 1992.
31. M. C. Bashaw, A. Aharoni, and L. Hesselink, "Phase–conjugate replay for *a*–axis strontium barium niobate single–crystal fibers," *Optics Letters* **18**(23), pp. 2059–2061, 1993.
32. F. Zhao and K. Sayano, "Compact read–only memory with lensless phase–conjugate holograms," *Optics Letters* **21**(16), pp. 1295–1297, 1996.
33. J. J. P. Drolet, E. Chuang, G. Barbastathis, and D. Psaltis, "Compact, integrated dynamic holographic memory with refreshed holograms," *Optics Letters* **22**(8), pp. 552–554, 1997.
34. J. Feinberg, "Self–pumped, continuous–wave phase conjugator using internal reflection," *Optics Letters* **7**(10), pp. 486–488, 1982.
35. G. W. Burr and R. M. Shelby, "Pixel–matched phase–conjugate holographic data storage," *SPIE Optical Processing & Computing newsletter* , p. 8, Jan. 1999.
36. S. Ducharme, J. C. Scott, R. J. Twieg, and W. E. Moerner, "Photorefractive behavior in a polymer?," *Physics Review Letters* **66**, p. 1846, 1991.
37. E. Mecher, R. Bittner, C. Brauchle, and K. Meerholz, "Optimization of the recording scheme for fast holographic response in photorefractive polymers," *Synthetic Metals* **102**(1–3), pp. 993–996, 1999.
38. N. Hampp, C. Brauchle, and D. Oesterhelt, "Bacteriorhodopsin wildtype and variant aspartate–96 → asparagine as reversible holographic media," *Biophysics Journal* **58**, pp. 83–93, 1990.
39. 7DCST:PCBM:DPP:PVK in the ratio 25%:0.5%:29%:45.5%, 53μm thick, 53V/μm, supplied by Erwin Mecher, Francisco Gallego, and Klaus Meerholz, Ludwig Maximilians Universitat Munchen, Institut fur Physikalische Chemie, 81377 Munchen, Germany.
40. M–type holograms in $\sim$80μm thick material optimized for low light optical interferometry, supplied by Thorsten Juchem and Norbert Hampp, Institut fur Physikalische Chemie, Universitat Marburg D–35032 Marburg, Germany.