SPIE Conference on *Critical Technologies for the Future of Computing*
August 2, 2000 Paper **4109–30**

Content-addressable holographic databases

Felix Grawert, Sebastian Kobras, Geoffrey W. Burr[1], Hans Coufal,
Holger Hanssen, Marc Riedel, C. Michael Jefferson, and Mark Jurich

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

1. ABSTRACT

Holographic data storage allows the simultaneous search of an entire database by performing multiple optical correlations between stored data pages and a search argument [1,2]. We have recently developed fuzzy encoding techniques for this fast parallel search and demonstrated a holographic data storage system that searches digital data records with high fidelity [2]. This content–addressable retrieval is based on the ability to take the two–dimensional inner product between the search page and each stored data page. We show that this ability is lost when the correlator is defocussed to avoid material oversaturation [3], but can be regained by the combination of a random phase mask and beam confinement through total internal reflection. Finally, we propose an architecture in which spatially multiplexed holograms are distributed along the path of the search beam, allowing parallel search of large databases.

Keywords: content–addressable data storage, volume holographic data storage, optical correlation, parallel search

2. INTRODUCTION TO HOLOGRAPHIC DATA STORAGE

Holographic data storage offers the benefits of large storage densities, fast parallel access, and rapid searches in large databases [4–9]. 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.

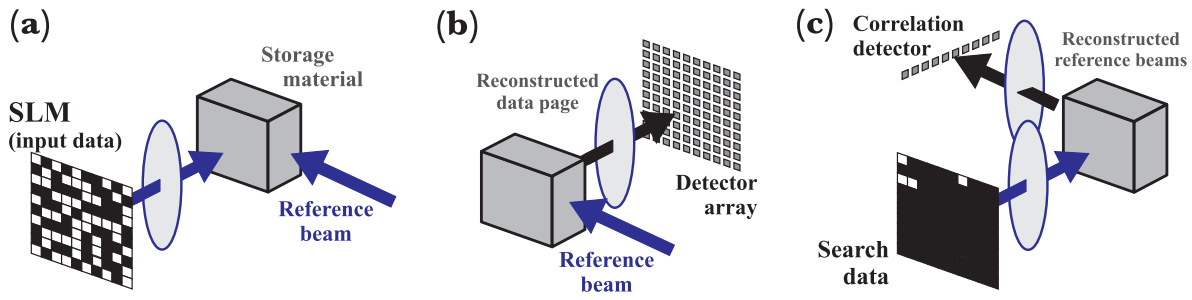

Figure 1: Holographic data storage system. **(a)** Two coherent beams, one carrying a spatial page of information, interfere within a photosensitive material to record a hologram. **(b)** Illuminating the hologram with the reference beam reconstructs a weak copy of the original information–bearing beam 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] To contact G. W. Burr, Email: *burr@almaden.ibm.com*; Tel: (408) 927–1512; Fax: (408) 927–2100.

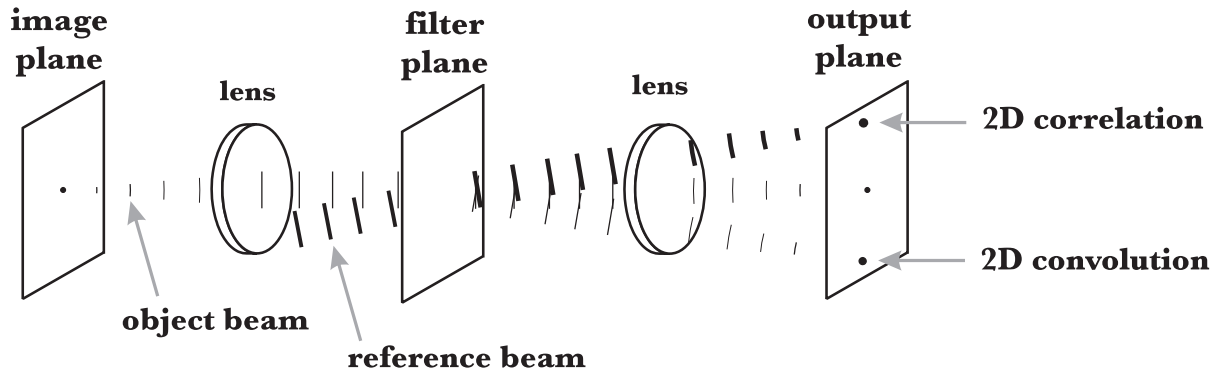

Figure 2: 4–F system, composed of two identical lenses separated by the sum of their focal lengths. Because each lens performs a spatial Fourier transform, the system performs both the 2–D correlation and the 2–D convolution between the pair of 2–D optically–input functions.

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 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 [10]. 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 [11].

3. 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, 9]. The amount of power diffracted into each output beam is proportional to the correlation between the input data page and the stored data page. The ability to perform a correlation is based on the property that a lens takes a two–dimensional (2–D) Fourier transform (FT) in spatial coordinates [12]. The spatial pattern in the back focal plane of any lens is the 2–D Fourier transform of the spatial pattern presented in its front focal plane. This property is used by holographic data storage systems in the conventional 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 [13], 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 [12]. 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 (see Figure 3) [12,14]. This method is widely used for the correlation of analog pages, such as in biometrics and target recognition [14–17], 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

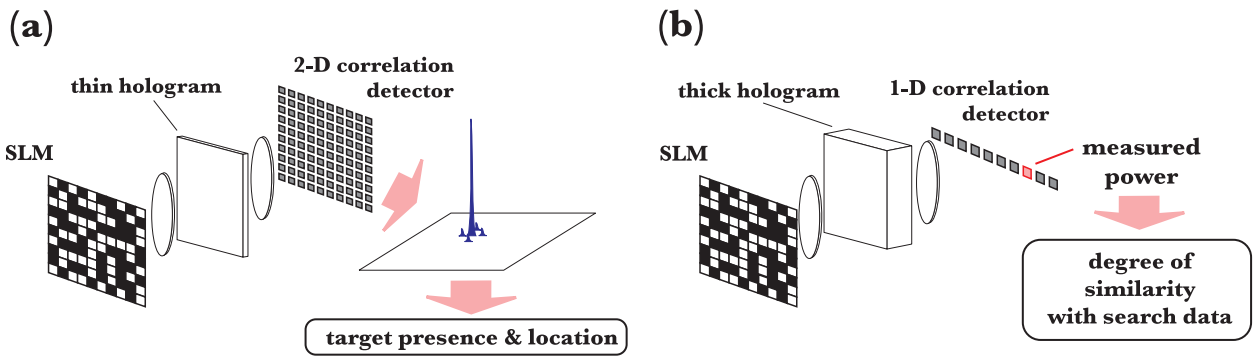

Figure 3: Correlation functions obtained in thin and thick holograms. **(a)** Thin holograms yield the entire 2–D correlation function, identifying the presence and location(s) of a sub-image within the input page. One example application is target recognition, where the sub-image of an airplane might be stored as a hologram for real–time identification within a large input page. **(b)** In thick holograms, Bragg mismatch limits the output to only a narrow, one-dimensional slice of the 2–D correlation function. The intensity in each correlation peak is an analog measure of the similarity between the associated stored page and the search argument. If each hologram represents a data record, the entire database can be compared against the search argument simultaneously.

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 [14]. 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, 9].

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. Therefore, in a digital holographic database, system performance depends on:

- collecting sufficient signal power in the integration time given to overcome detector and other noise sources [3],
- suppressing spurious signal contributions such as scatter, crosstalk from other holograms, and signal due to the OFF pixels of the SLM [3, 18],
- reducing hologram-to-hologram variations such as changes in diffraction efficiency [3, 18], and
- the accuracy with which the detected signal power measures the 2–D inner product [3]. It is crucial that the detected signal be proportional to the similarity between the search and stored data pages. As we will show below, this is not always the case.

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.

4. DATA ENCODING FOR HOLOGRAPHIC SEARCH

For 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 are 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 [1–3].

In addition to errors caused by weak signal, errors can arise from any signal unintentionally added to the correlation peak by cross–correlation between pixels on the SLM. Fortunately, since Bragg-mismatch eliminates the horizontal shift-invariance, pixels in one column of the SLM cannot produce a diffracted signal with the portion of the hologram written by pixels from any other column. Thus, different columns can be treated independently and neighboring data fields may be closely spaced horizontally. However, the vertical shift-invariance is conserved and therefore energy from ON-pixels in one row may be diffracted to the correlation plane if a stored page had pixels of neighboring rows turned ON. This results in completely misleading correlation peaks which are only slightly offset from the real one (representing the 2–D inner product). To avoid this, several rows of the SLM often have to be left unused between neighboring data fields, leading to less stored information per page.

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. Gray codes can be used to guarantee that neighboring data values retain similar encodings, but they do not prevent disparate values from having equally similar encodings.

4.1 FUZZY DATA ENCODING FOR FINDING SIMILAR RECORDS

We have developed a novel data-encoding method which allows similarity or fuzzy [19] searching, by encoding similar data values into similar pixel block patterns [2]. As shown in Figure 4, 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, the square root of the detector signal has to be taken. Here we see that the detected intensity faithfully follows the square of the 2–D inner product between the search and stored data pages, resulting in an accurate measure of similarity.

With this method, the width and height of the slider determine the output power at the detector and the range of the search, respectively. This enables a large variety of search operations: Not only is it possible to search for the closest match of the search argument ('='-operation; using sliders of same height in the stored and

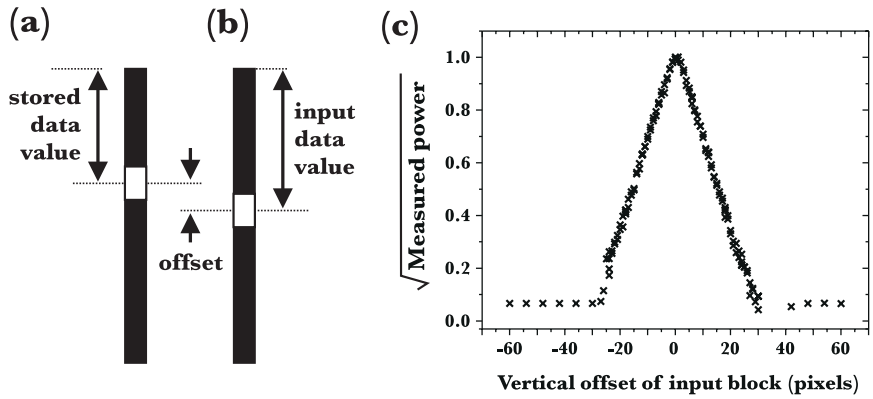

Figure 4: Data encoding for fuzzy searching. **(a)** When storing a hologram, a small block of SLM pixels are turned ON at some location within a predefined rectangular portion ('slider' track) of the data page. **(b)** For correlation readout, an input query is encoded as a similar block within the *same* track. **(c)** Any offset between the two blocks causes the brightness of the correlation peak to decrease. By encoding data values with the center position of the pixel block, the holographic system can now measure the similarity between data records and the input query, implementing fuzzy searching.

searching page), but it is also possible to find all values larger than a threshold ('>'-operation; by turning all pixels beyond the threshold in the search page ON) or to search in multiple ranges ('OR'-operation, by using two slider blocks in the slider track). Thus, the analog nature of the optical output becomes an enabling feature instead of a drawback [2].

If the cost in SLM pixels, which is proportional to the highest encoded data value, becomes unacceptably high, data fields may be made only partly fuzzy [2]. For example, in binary encoding, the least significant bits might be encoded with a smaller fuzzy slider, while exact encoding could be used for the most significant bits. This trades off search flexibility and reliability for more efficient use of SLM area. A holographic database might contain a blend of fuzzy, partly fuzzy and exact-coded data fields, depending on the degree of similarity matching required [2].

4.2 SEARCH FIDELITY AND ERROR PREVENTION

Because of the analog nature of the optical correlation process, low-overhead digital error correction is not available for the parallel search operation. To increase signal strength and thus improve the signal to noise ratio, one could assign a larger number of pixels to each field of the database, but at the cost of a loss in capacity. In contrast, it is far more attractive to use the analog holographic database as a fast front-end filter to a sequential digital search engine [2]. This would conserve the advantages of the holographic storage system—speed and capacity—while satisfying the constraint of low error. For example, the system might be asked to find the 10 best matches in a database with 1 million 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, 9]. 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 10,000.

4.3 EXPERIMENTAL HOLOGRAPHIC DATABASE DEMONSTRATION

To demonstrate high-fidelity parallel searching of a holographic content-addressable memory, we stored a small multimedia database [2] in the DEMON I system [20]. Each hologram represented one record from an IBM query-by-image-content (QBIC) database. In the QBIC system [21], searches are performed across feature vectors previously extracted from the images, rather than on the images themselves. Each record included several alphanumeric fields (such as image description and image number) encoded for exact searches, and 64 fuzzy sliders containing the color histogram information (percentage of each given color within the associated

image). A separate portion of the SLM page, pixel-matched onto a CCD detector for conventional address-based holographic readout, was encoded with the binary data for the small color image. One hundred holograms were recorded in a 90-degree-geometry $LiNbO_3$ crystal, with the reference angles chosen so that each reference beam was focused to a unique portion of the correlation camera [2].

Each search, initiated by a user query, ran under computer control, including display of the appropriate patterns, detection of the correlation peaks (averaging eight successive measurements to reduce detector noise), calibration by hologram strength, identification of the eight highest correlation scores, mapping of correlation bins to reference-beam angle, address-based recall of those eight holograms, decoding of the pixel-matched data pages, and finally, display of the binary images on the computer monitor. The optical readout portion occupied only 0.25s of the total 5s cycle [2]. To find images based on color similarity, 64 sliders were used to input the color histogram information for the images into the holographic storage system. The sliders could also be used to select images by color distribution, e.g., to search for images containing ∼20% white and ∼20% gray. We achieved high fidelity with fuzzy searches for similarly colored images, fuzzy searches for specific color percentages, and with exact searches on keyword [2].

5. AMBIGUOUS BEHAVIOR OF THE CORRELATION SCORE

When using correlation for target recognition, the position and sharpness of the correlation peak (relative to the background noise) determine system performance [14]. The content addressable database, however, relies upon the accuracy with which the power in the correlation peak measures the similarity between the search page and the stored data pages. Without this, the ranking of data pages becomes more ambiguous, and the system often overlooks the best match in favor of a less–similar but higher–scoring page. Fortunately, the holographic system shown in Figure 2 implements a true correlation, independent of the spatial patterns being compared. Given that the system uses coherent light, we need only take the square root of the measured power of the correlation peak to get the 2–D inner product between the search and stored data pages.

Unfortunately, we cannot use the system shown in Figure 2 for a fairly minor reason: The Fourier transform of a purely amplitude-modulated object beam contains a high intensity "DC" component. This object beam component focusses in the back focal plane of the imaging lens, reaching intensities that easily exceed the dynamic range of the photosensitive material. In order to avoid the resulting loss in reconstruction fidelity of the hologram, the correlator must typically be 'defocussed,' by placing the storage material a few centimeters in front of or behind the back focal plane [22]. Consequently, it is not exactly the Fourier transform that is stored in the storage material, but the Fresnel transform, and the correlator no longer truly performs correlations. On the one hand, this can be an advantage: it has been shown that the defocus of a correlator reduces the vertical shift invariance of the system [22]. This allows correlation peaks to be placed closer together on the output plane (without losing track of which peak is which), increasing the number of different sub–images a target recognition system can track. On the other hand, as we show here, with the defocussed correlator the relationship between the detected power in the correlation peak and the 2–D inner product depends on the particular patterns being compared.

The serious consequences of this ambiguity can be illustrated in a brief example: In Figure 5 a search page and three different stored data pages are shown, with a small number of blocks of pixels turned ON in each. The search page and the first stored data page have one block in common, and we normalize the resulting correlation score to one. The second page has two blocks, separated in the x-direction, in common with the search page. As expected, the coherent nature of the readout light causes the correlation peak to have four times as much power. By taking the square root, the proper 2–D inner product can be obtained. Following this logic, the third stored page should produce a correlation peak with nine times as much power as the single block produced. However, when one performs this experiment in the laboratory, this pattern results in a normalized score of *three*, and the measured signal now follows the 2–D inner product *linearly* rather than quadratically. As a result, the system will (incorrectly) find this third page to be less similar than the second. Thus, the defocussed correlator no longer provides accurate measurements of the similarity between search and stored data pages.

The origin of this behavior can be explained by looking at vertical and horizontal slices through the defocussed correlator. In Figure 6(a), the object beam is shown passing through the storage material on its way to the Fourier transform plane. Two identical blocks of pixels, separated by a large distance in the vertical direction, are displayed on the SLM. At the Fourier transform plane the optical fields from these two blocks would perfectly

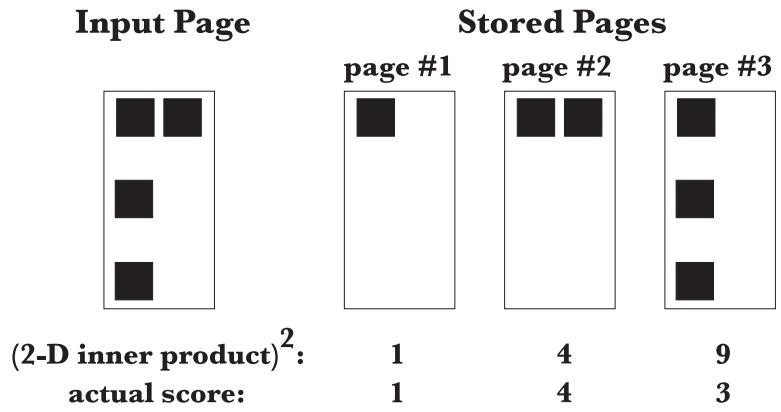

Figure 5: Example of the pattern–dependent behavior of the correlation score in a defocussed correlator. The relationship between correlation score and the 2D-inner product can vary from linear to quadratic, depending on the vertical spacing between blocks. This can lead to inaccurate ranking of similarity.

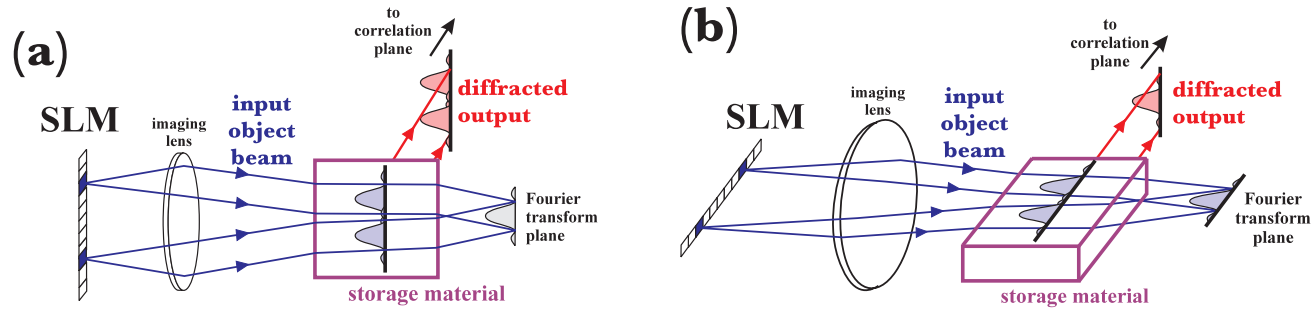

Figure 6: Origin of the difference between vertical and horizontal dependence of correlation score in the defocussed correlator. (a) Vertical slice through the optical system: the diffracted contributions from two vertically separated SLM pixels do not overlap, leading to a linear contribution toward total output power (see text). (b) Horizontal slice through the optical system: the optical fields from two horizontally separated SLM pixels are spatially distinct at the storage material, yet their contributions to the diffracted wavefront overlap completely. Thus the output power follows the square of the 2–D inner product.

overlap. However, in the defocussed correlator, the storage material is no longer at the Fourier plane. At the new position of the storage material, the fields from the two pixel blocks do not completely overlap, with their two mainlobes separated by a vertical distance. To perform an autocorrelation, we store a hologram of this composite object beam with a large reference beam, and then illuminate with the same object beam. The diffracted wavefront builds up horizontally along the x-axis, resulting in a reconstructed reference beam with the same vertical profile as the illuminating object beam. Adding the upper block to the lower thus increases the extent of the reconstructed reference beam, but does not increase its amplitude. Thus the total diffracted power doubles (twice the area $\times$ same output intensity), and the resulting correlation signal follows the 2–D inner product *linearly*. In effect, the two object beam field components U appear to add "in intensity" ($S \propto (U^2 + U^2)$) to produce the correlation power S, even though the light is still spatially and temporally coherent [3].

In contrast, the fields of closely spaced pixel blocks overlap even in the defocussed storage material. Adding the upper block to the lower now doubles the amplitude of the reconstructed reference beam (without increasing its extent). Thus the total diffracted power quadruples (same area $\times$ four times the output intensity), and the correlation signal follows the *square* of the 2–D inner product. Here the two object beam field components do add "in field" ($S \propto (U + U)^2$) as would be expected for coherent light.

The result of this is that the system produces the 2–D inner product for blocks which are widely separated, but outputs the square of the 2–D inner product for blocks which are closely spaced. What's worse is that the system will smoothly vary between these two behaviors for intermediate spacings, as shown experimentally in Figure 7.

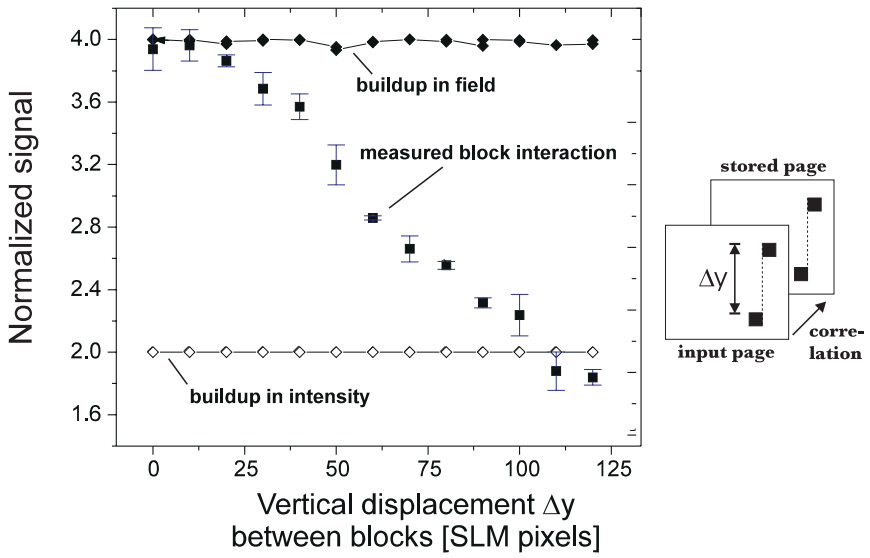

Figure 7: Dependence of normalized signal strength on the vertical separation between two pixel blocks. The aggregate signal (squares) is shown, as well as with the aggregate signal due to each block individually, shown added directly (open diamonds), and added in quadrature (filled diamonds).

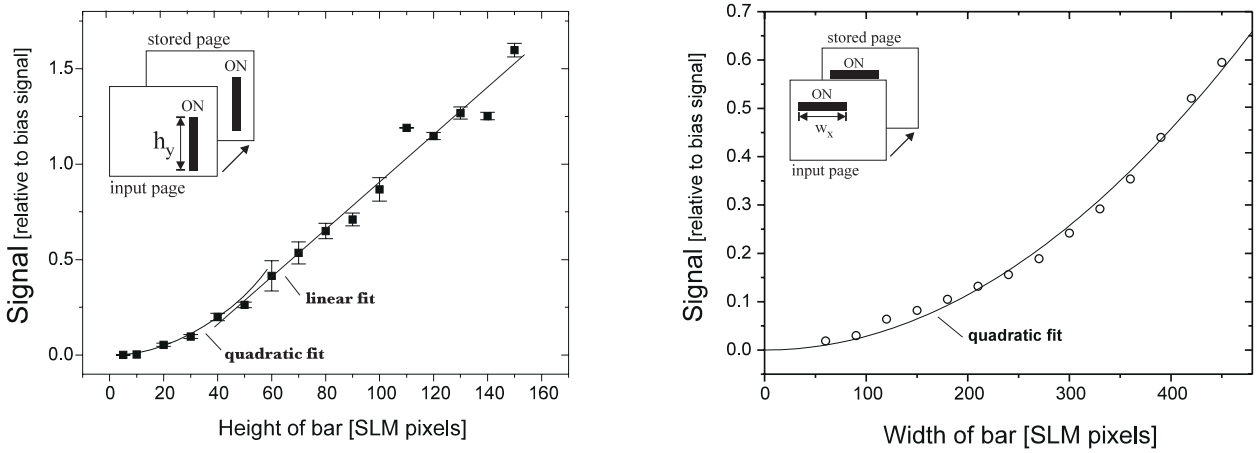

Figure 8: Dependence of correlation score on (a) the height of a pixel block, changing from quadratic to linear behavior; (b) the width of a pixel block, showing the expected quadratic dependence.

Here we plot the normalized autocorrelation score for two blocks of 15×15 pixels (middle curve, filled squares) as a function of their center-to-center separation in pixels [3]. The two blocks add 'in field' when they are close together, add 'in intensity' for large separations, and follow some intermediate behavior in between [3]. The upper and lower curves come from the two correlation peaks ($s_1 \sim s_2 \sim U^2$) produced by the blocks when presented individually. The top curve (filled diamonds) represents the signal expected if the two blocks were to always add "in field" ($S \propto (U+U)^2 = 4U^2$), derived from the measured data as $(\sqrt{s_1} + \sqrt{s_2})^2$. The bottom curve (open diamonds) shows the signal that would result if the two blocks always added "in intensity" ($S \propto U^2 + U^2 = 2U^2$), produced by $s_1 + s_2$. All three curves are normalized by forcing this lower curve to the constant value of 2.0.

This pattern–dependent behavior in the defocussed correlator also affects the dependence of the autocorrelation score of a single pixel block on its vertical size. For small blocks, the object field from all pixels overlaps at the storage material, and the correlation power grows quadratically. However, when the block is large and the field from its top and bottom pixels no longer overlap in the storage material, the signal strength grows linearly with

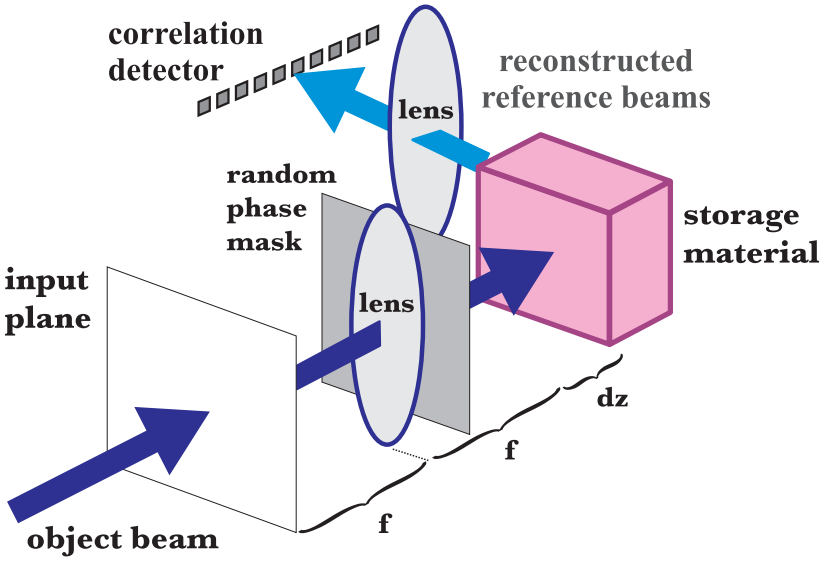

Figure 9: Modified holographic correlator system: A random phase mask (RPM) is placed shortly behind the first lens to distribute the optical power that would otherwise focus at the Fourier transform plane. The material is still defocussed away from the FT plane for the purposes of a with- and with-out RPM comparison.

block height. This behavior is shown experimentally in Figure 8(a).

This dependence of two–block signal on block separation does not show up in the horizontal direction, but only because we're using the 90° storage geometry. As shown in Figure 6(b), two widely separated blocks again produce spatially distinct field patterns in the defocussed storage material. However, since the diffracted beam builds up along the horizontal x axis, these various field contributions are merely integrated together to produce the same reconstructed reference beam. Their spatial distribution along x cannot affect the integrated amplitude. Thus, the power in the correlation peak represents the square of the 2–D inner product, just as if the fields had perfectly overlapped in a conventional correlator at the Fourier transform plane. Horizontally separated pixel blocks always 'add in field', and the correlation signal grows quadratically with the width of the block, as shown experimentally in Figure 8(b). Note that for a defocussed correlator using a transmission hologram, however, both horizontal and vertical block separations would experience this pattern–dependent correspondence between measured power and 2–D inner product.

6. RANDOM PHASE MASK AND SPATIAL MULTIPLEXING

To avoid the nonlinearity in the correlation function of the defocussed correlator, we propose the use of a pixelated random phase mask (RPM) placed between the imaging lens and the storage material (Figure 9). Consisting of pixels which add a random phase of either 0 or π to the incident light, the RPM introduces additional spatial bandwidth to the light incident on the storage material. Consequently, energy from a small pixel block that would otherwise concentrate in a narrow mainlobe is now distributed across the Fourier transform plane. Thus, the storage material can now be placed at the Fourier transform plane. Furthermore, even at positions a few centimeters away from the Fourier transform plane, the energy originating from small pixel blocks now overlaps enough to prevent the position–dependent performance described above.

In order to show this, we numerically simulate the optical system. Previously, we have done this by using an analytical expression for the impulse response of the system (from SLM to storage material), and then numerically integrating over the finite size of the pixel [3]. However, because of the random nature of the RPM, the proposed optical system shown in Figure 9 can not be treated analytically. Hence we have written a 2–D simulation in Matlab for propagating the optical fields through the system as complex matrices, implementing Fresnel diffraction by using the Fast Fourier Transform (FFT) algorithm to switch back and forth between spatial and angular spectrum representations. With the optical system used in Reference [3], the simulation shows excellent

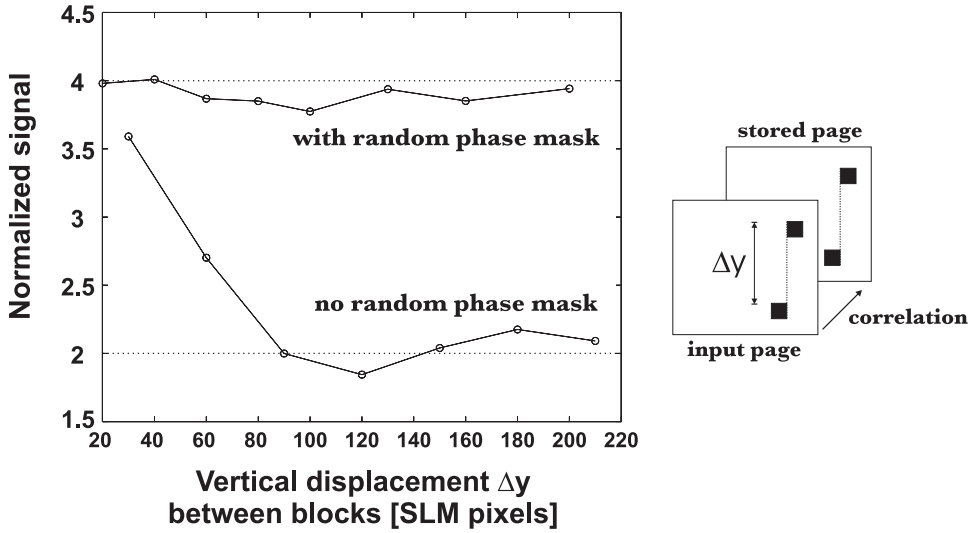

Figure 10: Simulated dependence of the normalized signal strength on the vertical separation between two pixel blocks. When the random phase mask is added, the nonlinearity disappears and the normalized correlation score remains at a value of four.

agreement with the measured data in Figure 7. With the optical system shown in Figure 9[1], the simulation shows that simply adding the random phase mask does indeed remove the pattern–dependent correspondence between correlation score and the 2–D inner product, as shown in Figure 10. The random phase mask forces the auto–correlation of a two block pattern to produce roughly four times as much signal as a single block. We also expect that the RPM will tend to reduce the remaining shift invariance, allowing the SLM to be tightly packed with data fields and the output correlation plane to be filled with sharp, well–defined correlation peaks.

Even with the random phase mask, however, the energy from widely separated pixel blocks on the SLM will tend to diverge as the object beam propagates away from the Fourier transform plane. This limits the range of positions around the Fourier plane that can be used for storage media placed along the path of the object beam, and thus the number of holograms that can be interrogated by a single optical pulse. We can reduce the absorption loss by using gated, two–color $LiNbO_3$ crystals (see Reference [23, 24] and references within), so that the red or infrared object beam is only absorbed when we use blue or ultraviolet gating light to excite intermediate trap levels in the crystal. However, there will still be a fairly low limit on the number of storage locations imposed by the RPM–modulated object beam, with its convergence towards and then divergence away from the Fourier transform plane.

A solution to this problem is to use a single, long crystal bar of fairly small cross–sectional aperture. As shown in Figure 11, once the object beam enters the end of the crystal bar (at or near the Fourier transform plane), the object beam will remain confined by the total internal reflection. This allows one to greatly increase the number of spatially–multiplexed storage locations, while retaining the parallelism of the associative search. Because speed of recording is not particularly important in this application, the reference beam can be brought to each spatial position by mechanical translation of a lens. Since each reference beam is already focussing when it is holographically recorded, the reconstructed beams diffracted by the illuminating search beam will focus onto the correlation plane detector without the need for any intervening optics. While the limit on the number of holograms per storage location remains the same (dominated by material dynamic range versus background noise), the number of locations is no longer limited by the geometry of the object beam. Instead, issues such as crystal fabrication and residual losses within the material will dominate.

With the random phase mask alone, the holographic database engine provides consistent similarity measure-

[1]System parameters are as follows: 514.5nm collimated illumination of SLM at front FT plane, 12.8μm pixels, imaging lens of focal length 50mm, RPM 10mm behind lens with random binary phase pixels of 9μm, storage material with 3×3mm^2 square aperture placed 10mm downstream from back FT plane, correlation lens of 50mm, correlation score derived from central five 9μm detector pixels.

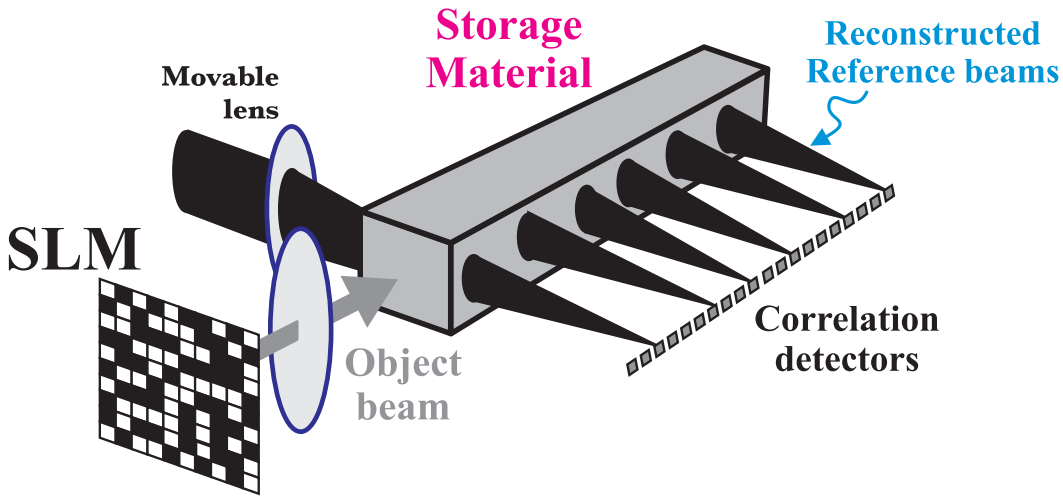

Figure 11: Holographic system for massively parallel content–addressable search. The object beam is confined by total internal reflection so that multiple storage locations may be simultaneously addressed by a single search page.

ments that are pattern–independent, but is limited in the number of accessible records. With beam confinement of the object beam alone, the holographic system can access many more records, but provides an ambiguous, pattern–dependent similarity metric. With *both* the random phase mask and total internal reflection, we can build a massively parallel system which accurately measures pattern similarity.

7. CONCLUSIONS

Because of its inherent parallelism, volume-holographic content addressable data storage is an attractive technique for searching large databases with complex queries. Fuzzy encoding techniques allow searches for records containing data values that are *near* those in the holographically stored records. This ability to identify matching data pages requires, however, that the power in the detected correlation peaks be an accurate measure of the similarity between search and stored data pages. We showed that with a defocussed correlator, the output signal power becomes pattern–dependent, and demonstrated how this can lead the holographic system to misrank the similarity of stored database records. We proposed that introducing a random phase mask into the object beam could solve this problem, and verified this with a numerical simulation. If, in addition, we confine the object beam by total internal reflection, we can combine angle and spatial multiplexing to bring this rapid parallel optical search to large databases. Future work includes completing a test platform to test the random phase mask and total internal reflection experimentally.

8. REFERENCES

[1] B. J. Goertzen and P. A. Mitkas. Volume holographic storage for large relational databases. *Optical Engineering*, 35(7):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):6779–6784, 1999.

[3] S. Kobras. Associative recall of digital data in volume holographic storage systems. Master's thesis, Technische Universität München, Munich, May 1998.

[4] D. Psaltis and F. Mok. Holographic memories. *Scientific American*, 273(5):70, 1995.

[5] J. H. Hong, I. McMichael, T. Y. Chang, W. Christian, and E. G. Paek. Volume holographic memory systems: techniques and architectures. *Optical Engineering*, 34:2193–2203, 1995.

[6] J. F. Heanue, M. C. Bashaw, and L. Hesselink. Volume holographic storage and retrieval of digital data. *Science*, 265:749, 1994.

[7] D. Psaltis and G. W. Burr. Holographic data storage. *IEEE Computer*, 31(2):52–60, 1998.

[8] 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 Journal of Research and Development*, 44(3):341–368, May 2000.

[9] G. W. Burr and P. Mitkas. *Holographic Data Storage*, chapter Volume holographic correlators. Springer–Verlag, Berlin, 2000. ed., H. Coufal and D. Psaltis and G. Sincerbox.

[10] G. W. Burr. *Volume holographic storage using the 90° geometry*. PhD thesis, California Institute of Technology, Pasadena, Calif., 1996.

[11] R. M. Shelby, J. A. Hoffnagle, G. W. Burr, C. M. Jefferson, M.-P. Bernal, H. Coufal, R. K. Grygier, H. Günther, R. M. Macfarlane, and G. T. Sincerbox. Pixel–matched holographic data storage with megabit pages. *Optics Letters*, 22(19):1509–1511, 1997.

[12] J. W. Goodman. *Introduction to Fourier Optics*. McGraw–Hill, New York, 1968.

[13] A. Vander Lugt. Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory*, IT–10:139–145, 1964.

[14] S. H. Lee, editor. *Optical Information Processing–Fundamentals*. Springer–Verlag, Berlin, 1981.

[15] E. G. Paek and D. Psaltis. Optical associative memory using Fourier–transform holograms. *Optical Engineering*, 26(5):428–433, 1987.

[16] H.-Y. S. Li, Y. Qiao, and D. Psaltis. Optical network for real time face recognition. *Applied Optics*, 32(26):5026–5035, 1993.

[17] A. Pu, R. Denkewalter, and D. Psaltis. Real time vehicle navigation using a holographic memory. *Optical Engineering*, 36(10):2737–2746, 1997.

[18] G. A. Betzos, A. Laisne, and P. A. Mitkas. Improved associative recall of binary data in volume holographic memories. *Optics Communication*, 171(1–3):37–44, 1999.

[19] L. A. Zadeh. Fuzzy sets. *Information control*, 8:338–353, 1965.

[20] 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):639–641, 1997.

[21] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, B. Qian Huang Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

[22] M. Levene, G. J. Steckman, and D. Psaltis. A method to control the shift invariance of optical correlators. *Applied Optics*, 38(2):394–398, 1999.

[23] H. Guenther, R. M. Macfarlane, Y. Furukawa, K. Kitamura, and R. R. Neurgaonkar. Two-color holography in reduced near-stoichiometric lithium niobate. *Applied Optics*, 37(32):7611–7623, 1998.

[24] K. Buse, A. Adibi, and D. Psaltis. Non volatile holographic storage in doubly doped lithium niobate crystals. *Nature*, 393(6686):665–668, 1998.