SPIE Conference on *Three– and Four–Dimensional Optical Data Storage*
July 31, 2001 Paper **4459–53**

Using volume holograms to search digital databases

Geoffrey W. Burr, George Maltezos, Felix Grawert,
Sebastian Kobras, Holger Hanssen, and Hans Coufal

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

ABSTRACT

Holographic data storage 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 have previously developed fuzzy encoding techniques for this fast parallel search, and holographically searched a small database with high fidelity [2]. We recently showed that such systems can be configured to produce true inner–products, and proposed an architecture in which massively–parallel searches could be implemented [3]. However, the speed advantage over conventional electronic search provided by parallelism [2] brings with it the possibility of erroneous search results, since these analog correlation scores are subject to various noise sources. We show that the fidelity of such an optical search depends not only on the usual holographic storage signal-to-noise factors (such as readout power, diffraction efficiency, and readout speed), but also on the particular database query being made. In effect, the presence of non–matching database records with nearly the same correlation score as the targeted matching records reduces the speed advantage of the parallel search. Thus for any given fidelity target, the performance improvement offered by a content–addressable holographic storage can vary from query to query even within the same database.

Keywords: content–addressable data storage, volume holographic 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 [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.

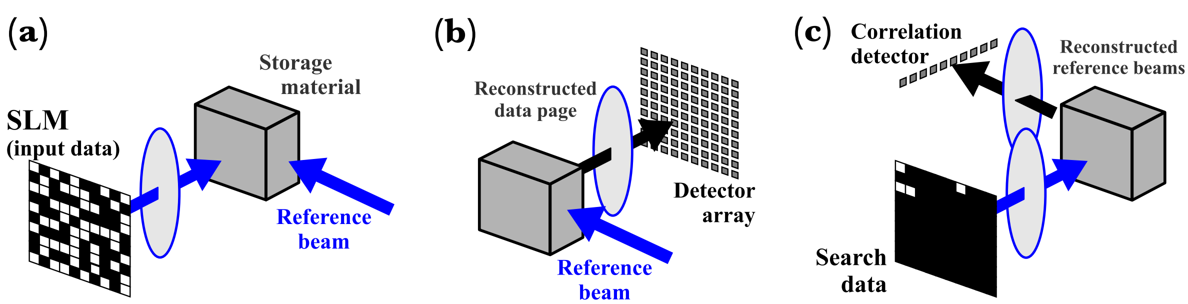


Figure 1. Holographic data storage system. **(a)** Two coherent beams, one carrying a spatial page of information, interfere within a photosensitive material to record a hologram. **(b)** Illuminating the hologram with the reference beam reconstructs a weak copy of the original information–bearing beam 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] For further information, contact G. W. Burr at *burr@almaden.ibm.com*.

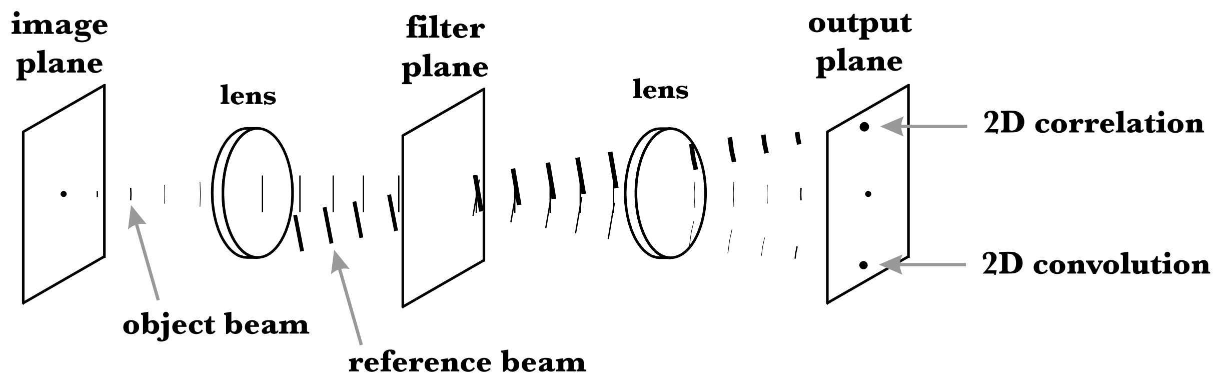


Figure 2. 4–F system, composed of two identical lenses separated by the sum of their focal lengths. Because each lens performs a spatial Fourier transform, the system performs both the 2–D correlation and the 2–D convolution between the pair of 2–D optically–input functions.

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].

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, 9]. 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 [12]. 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 [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 [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 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. Here, the system performance depends on:

- collecting sufficient signal power in the integration time given to overcome detector and other noise sources [18],
- suppressing spurious signal contributions such as scatter, crosstalk from other holograms, and signal due to the OFF pixels of the SLM [18, 19],
- reducing hologram-to-hologram variations such as changes in diffraction efficiency [18, 19], and
- the accuracy with which the detected signal power measures the 2–D inner product [18]. 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, 18].

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 [20] 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, 18], 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, 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 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'.

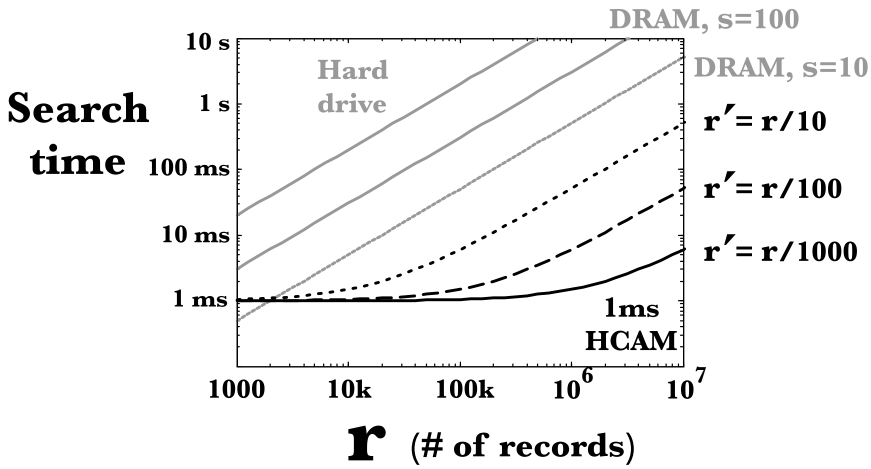


Figure 3. Plot of time required to search an entire database versus the number of database records, r. Each record is assumed to contain 1000 one–byte fields, of which s fields are to be extracted, weighted, and summed to compute the search metric. Gray lines for conventional technologies (Hard drive, 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. In the remainder of this paper, we describe 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.

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

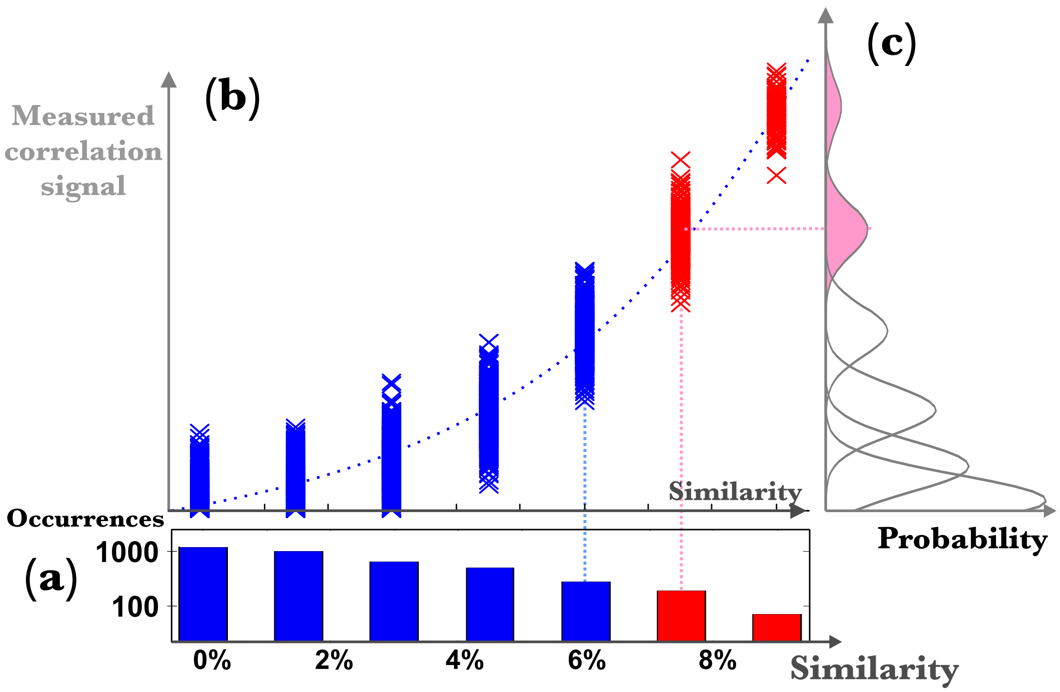


Figure 4. Relationship between expected score (the similarity between database records and the search query in terms of SLM area in common) and the actual score measured by the holographic search engine in the presence of noise (in detected photons). Part(a) describes the database and query in terms 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.

approximation applies) [3, 21], the relationship between the expected score in SLM pixels and the measured score in 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.

A given database search (the combination of the database and the particular query) can be described through a set of expected scores, $x_1, x_2, \ldots$, and the occupancy or frequency of each of these scores, $w_1, w_2, \ldots$. These functions describe the position and height of the bars in Figure 4(a). The sum of all w_i terms is the number of records in the database,

$$r = \sum_{\forall i} w_i. \tag{4}$$

Each expected score x_i (in units of common SLM area) is associated with an average measured score y_i (in units of photons). Applying the threshold y_t should ideally select all the "matching" records, which comprise some subset $\mathbf{x}'$

of $\mathbf{x}$. This subset may be in many cases simply those records with large expected value, or $\mathbf{x}' = x_i$ such that $i \geq j$, where x_j is the smallest expected score still considered to be a match.

If $p(y)$ represents the zero–mean probability density function of the random noise, then

$$P_1 \;=\; \int_{-\infty}^{y_t - y_i} p(y)\, dy \tag{5}$$

describes the probability that any one member x_i of $\mathbf{x}$ will have a measured signal that falls below y_t. If y_t is significantly lower than y_i, then the upper limit of the integral is well below zero and the overall probability should be low. The probability that all of the w_i database records with this same expected score exceed the threshold y_t (and thus are *not* erroneously excluded) is then

$$P_2 \;=\; \prod_{k=1}^{w_i} (1 - P_1). \tag{6}$$

Including all of the possible matching records involves simply accumulating the probability from all records in set x_i', or

$$P_3 \;=\; \prod_{\forall\, i \in \mathbf{x}'} \left[\prod_{k=1}^{w_i} (1 - P_1) \right]. \tag{7}$$

The probability of error, P_E, defined as the probability that one (or more) of the matching records fails to be selected by the cutoff threshold, y_t, is then simply $1 - P_3$.

One common noise source in holographic data storage is thermal noise in the detection electronics, which can become a dominant source of noise when signal levels become small [22]. The PDF $p(y)$ is then written as a zero–mean Gaussian,

$$p(y) = \frac{1}{\sqrt{2\,\pi}\sigma_e}\ \exp(-\frac{y^2}{2\,\sigma_e^2}), \tag{8}$$

and the overall probability of error becomes

$$P_e \;=\; 1 - \prod_{\forall\, i \in \mathbf{x}} \left[\prod_{k=1}^{w_i} \left(1 \;-\; \frac{1}{2}\,\mathrm{erfc}\left(\frac{y_i - y_t}{\sqrt{2}\,\sigma_e} \right) \right) \right], \tag{9}$$

where σ_e is the standard deviation of the thermal noise, in equivalent noise photons (e.g., referenced through the quantum efficiency of the detector).

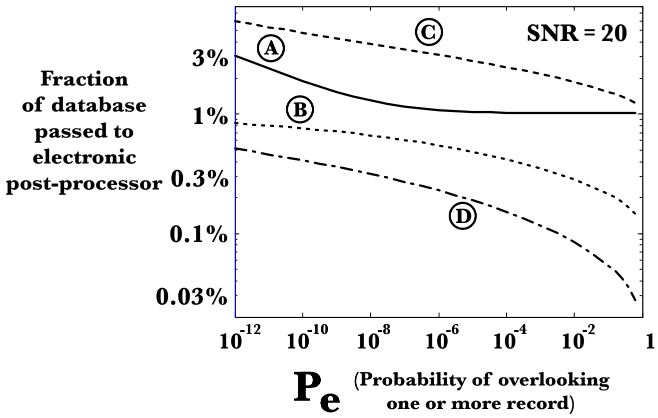
The number of records passed to the electronic system to be checked is then the number of correct matching records plus the overhead (the number of superfluous near–matching records), written as

$$r_m + r_{ov} \;=\; \sum_{\forall\, i \in \mathbf{x}'} w_i \;+\; \sum_{\forall\, i \notin \mathbf{x}'} \left[\frac{w_i}{2}\,\mathrm{erfc}\left(\frac{y_t - y_i}{\sqrt{2}\,\sigma_e} \right) \right], \tag{10}$$

As y_t is reduced, P_e is also brought down but only at the cost of increased overhead.

In Figure 5, we plot the percentage of the database passed to the electronic post–processor as a function of the probability of making an erroneous database search, (P_e), for a number of different database searches. These plots assume an SNR of 20, defined as the ratio between the maximum signal, $y_{i_{max}}$, and the standard deviation of the Gaussian noise, σ_e. The four database searches are defined in Figure 7, which shows the distribution of expected signals (w_i vs. x_i) and identifies which records are considered to be matching. As expected, 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. Note that search **A**, in which the matching records are well–separated in expected score from non–matching records, is a particularly easy search to perform. Very few additional records need to be passed until the desired P_e becomes quite low. However, as shown by search **B**, simply redefining the



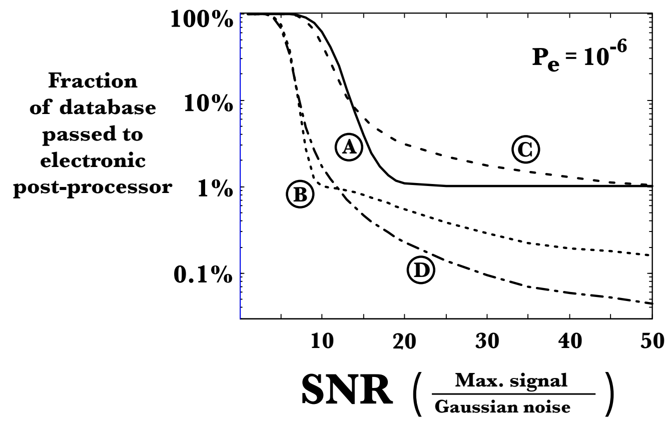


Figure 5. The percentage of the database passed to the electronic post–processor, as a function of the probability, P_e, of an erroneous search (one or more matching records are overlooked and not passed along). The SNR, or ratio between maximum signal and the standard deviation of the Gaussian noise, is fixed at 20.

Figure 6. The percentage of the database passed to the electronic post–processor, as a function of the SNR (the ratio between maximum signal and the standard deviation of the Gaussian noise). The probability, P_e, of an erroneous search is fixed at 10^{-6}.

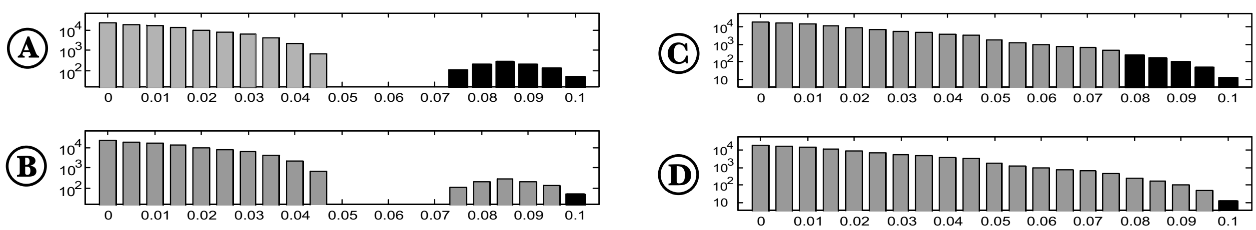


Figure 7. The four example database searches used, showing occurrences on a log scale as a function of the expected similarity. Black bars represent matching expected scores, while gray bars represent the range of scores for non–matching records. The total number of records r is 100,000 for each. Note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

subset of matching records makes the search much more difficult. In essence, 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.

Figure 6 also plots the same metric (percentage of the database passed to the electronics), but as a function of the SNR, at a constant P_e of 10^{-6}. Note that when the SNR gets too low, the holographic search engine is essentially useless, passing on almost all of the records. This plot can be used to connect the P_e and fraction of database passed to the other performance parameters of the holographic system, such as number of holograms, number of database fields per holographic data page, and optical search speed.

7. RELATIONSHIP BETWEEN SNR AND SYSTEM PERFORMANCE METRICS

The SNR as described in the previous section is simply the ratio between the maximum correlation signal for a particular search and the standard deviation of the Gaussian noise. The latter can easily be related to an experimental setup: it is the number of noise electrons at each camera pixel. Computing the number of signal electrons is a bit more involved.

The power illuminating the storage material in the object beam during the recording of a datapage (or the subsequent autocorrelation of this page) can be written as

$$P_{in,\,auto} \;\equiv\; P_S\,\frac{N_r}{N_S}\,\eta_o \;=\; \left(\sum_{i=1}^{N_r} S_i\right)^2 \;=\; N_r\,\left|S\right|^2\,, \tag{11}$$

where P_S is the power illuminating the SLM, N_S is the number of pixels in the SLM, N_r is the number of pixels turned ON to encode this database record, η_o is the optical efficiency between SLM and storage material (including the effects of higher SLM orders and the dead space between SLM pixels), and S_i is the field amplitude at the storage material associated with pixel i. Since this spatial frequency component originated at a single pixel on the SLM, the optical power per pixel must be

$$|S|^2 \;=\; \frac{P_S}{N_S}\,\eta_o. \tag{12}$$

Similarly, the optical power illuminating the storage material during a search query is

$$P_{in,\,query} \;\equiv\; P_S\,\frac{N_q}{N_S}\,\eta_o \;=\; \left(\sum_{i=1}^{N_q} S_i\right)^2 \;=\; N_q\;|S|^2 \tag{13}$$

where the major difference is the smaller number of pixels N_q. The correlation aspect of this operation is included by defining N_q as the number of pixels *in common* between search query and stored datapage. Thus this optical power is the total illuminating power that will lead to significant holographic reconstruction, and combination of high SLM pixel contrast, Bragg selectivity, and the use of random phase masks [3] is assumed to suppress all other holographic contributions.

The stored hologram between search page and reference beam R is written as

$$\sum_{i=1}^{N_r} S_i^*\,R, \tag{14}$$

where the spatial frequency components are ignored for simplicity. Reconstructing that hologram with the same object beam removes the object–derived spatial frequency components, leaving a reconstructed reference beam with optical power

$$P_{out,\,auto} \;=\; \left|\sum_{i=1}^{N_r} S_i\;S_i^*\;R\,\sqrt{\eta_p}\right|^2 \;\equiv\; P_{in,\,query}\,\eta_h \;=\; N_r\;|S|^2\left(\frac{M/\#}{M}\right)^2, \tag{15}$$

where η_p is the diffraction efficiency of the ith pixel component, and η_h is the diffraction efficiency of the autocorrelation reconstruction as a whole. The fourth term of Equation 11 has been substituted in order to include $|S|^2$. We adopt the conventional representation for the scaling of this diffraction efficiency with the number of superimposed holograms M, with a single scaling coefficient termed $M/\#$ [23], so that

$$\eta_h \;\equiv\; \left(\frac{M/\#}{M}\right)^2. \tag{16}$$

Solving for the diffraction efficiency of each pixel component in terms of the entire hologram's efficiency, we find

$$\eta_p = \left(\frac{M/\#}{M}\right)^2\;\frac{1}{N_r\;|S|^2\;R^2}. \tag{17}$$

We can then adapt Equation 15 slightly to represent the diffraction efficiency of the search query as

$$P_{out,\,query} \;=\; \left|\sum_{i=1}^{N_q} S_i\;S_i^*\;R\,\sqrt{\eta_p}\right|^2 \;=\; P_S\,\frac{N_q^2}{N_S\,N_r}\,\eta_o\,\left(\frac{M/\#}{M}\right)^2, \tag{18}$$

where both Equations 12 and 17 have been substituted to remove η_p and then the remaining term of $|S|^2$.

This is then the output power in the reconstructed reference beam, and thus represents the signal power received by the correlation detector for this particular search query. As a check, we can set $N_q = N_r$ to solve for the output power for querying with the full database page, obtaining

$$P_{out,\,auto} \;=\; P_S\,\frac{N_r}{N_S}\,\eta_o\,\left(\frac{M/\#}{M}\right)^2, \tag{19}$$

which agrees with Equations 15 and 12.

We can redefine this power in terms of the *fraction* of the SLM turned ON,

$$f_r \equiv \frac{N_r}{N_S} \qquad f_q \equiv \frac{N_q}{N_S}. \tag{20}$$

The number of signal photoelectrons received by this correlation detector is

$$n_s = P_{out,\,query}\,\frac{\eta_e}{h\nu}\,t_o, \tag{21}$$

where h is Planck's constant, ν the frequency of light used, η_e the quantum efficiency of the photodetector, and t_o the optical exposure time for the search query. Substituting Equation 18, we obtain

$$n_s = \frac{P_S}{h\nu}\,f_r\left(\frac{f_q}{f_r}\right)^2 \eta_o \left(\frac{M/\#}{M}\right)^2 \eta_e\,t_o. \tag{22}$$

We wish to solve for the desired performance metrics of the holographic system: M, t_o, and N, the maximum number of search elements, in terms of all other variables. We can represent N as the ratio between the SLM fractions for full database page and minimum search query,

$$N \equiv \frac{f_r}{f_q}. \tag{23}$$

The number of signal electrons that must be maintained, n_s, can then be related to the number of detector noise electrons, n_d, and the minimum SNR that must be maintained,

$$n_s = \mathrm{SNR}\,n_d. \tag{24}$$

The final result for the dependence between the number of superimposed pages, M, the number of search elements per page, N, and the optical search time, t_o, is then

$$\boxed{\frac{MN}{\sqrt{t_o}} = M/\#\sqrt{\frac{P_S}{h\nu}\,\frac{\eta_o\,\eta_e\,f_r}{n_d\,\mathrm{SNR}}}} \tag{25}$$

We plot the relationship between M and N in Figure 8. Assumptions include a wavelength of 532nm, 75 thermal noise electrons, an SNR of 20 ($\sigma_e = 0.05$), $\eta_o = \eta_e = f_r = 0.35$), an $M/\#$ of 1, an optical exposure time of 1 millisecond, and 200mW of optical power illuminating the SLM. Note that N search elements does not necessarily imply a limit of N fields—one can encode $5N$ fields if one ensures that all search queries always involve at least 5 fields. The limit of M holograms applies to each stack of superimposed holograms, but multiple storage locations can be arranged along the object beam so that a much larger database r can still be illuminated with a single optical exposure from a single SLM [3].

8. REALIZABLE IMPROVEMENT IN SEARCH SPEED

Now that we have connected the parameters of the holographic database machine to its performance, we can see the particular variations in search time performance for our four example database searches. Here we assume that an operating point on Figure 8 has been selected for the holographic database engine, and the datapages have been encoded and holographically stored. However, there still remains a variation in the probability of error from search to search, depending on the boundary between the matching and near–matching records. 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 Dataset **A** in Figure 9 for three different P_e targets. 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 all four searches in Figure 10. 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.

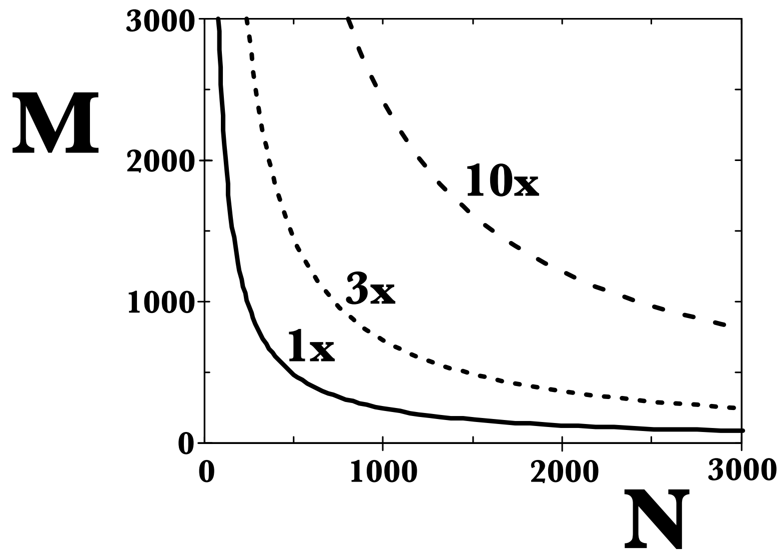


Figure 8. Number of superimposed database pages, M, versus the number of search elements per page, N, for the parameters listed in the text. The "3×" and "10×" curve imply an aggregate improvement in these parameters (the right side of Equation 25 plus the optical exposure time, t_o) by a factor of 3× and 10×, respectively.

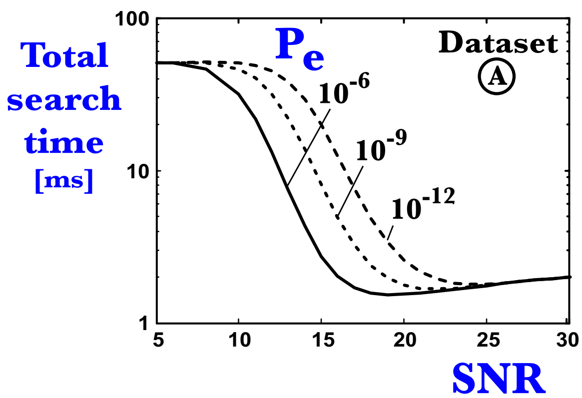


Figure 9. Plot of search time vs. SNR, showing the impact of lower SNR (many records must be passed to the electronic processor) and higher SNR (the optical exposure is longer than it needs to be). The ratio between the minimum and maximum search times represents the performance improvement offered by the holographic database engine.

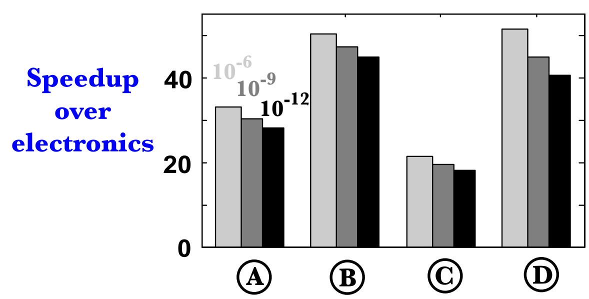


Figure 10. Factors by which the holographic database engine outperforms a conventional search, for the four database examples listed in Figure 7 and three different probability of error targets.

9. CONCLUSIONS

Because of 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}.

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 *Proceedings of the SPIE*, vol. 4109–30, 2000.
4. D. Psaltis and F. Mok, "Holographic memories," *Scientific American* **273**(5), p. 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**, pp. 2193–2203, 1995.
6. J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science* **265**, p. 749, 1994.
7. D. Psaltis and G. W. Burr, "Holographic data storage," *IEEE Computer* **31**(2), pp. 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 J. Research and Development* **44**, pp. 341–368, May 2000.
9. 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.
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. 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.
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**, pp. 139–145, 1964.
14. S. H. Lee, ed., *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), pp. 428–433, 1987.
16. H.-Y. S. Li, Y. Qiao, and D. Psaltis, "Optical network for real time face recognition," *Applied Optics* **32**(26), pp. 5026–5035, 1993.
17. A. Pu, R. Denkewalter, and D. Psaltis, "Real time vehicle navigation using a holographic memory," *Optical Engineering* **36**(10), pp. 2737–2746, 1997.
18. S. Kobras, "Associative recall of digital data in volume holographic storage systems," Master's thesis, Technische Universität München, Munich, May 1998.
19. 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.
20. L. A. Zadeh, "Fuzzy sets," *Information control* **8**, pp. 338–353, 1965.
21. H. J. Coufal, D. Psaltis, and G. Sincerbox, eds., *Holographic Data Storage*, Springer–Verlag, 2000.
22. 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.
23. F. H. Mok, G. W. Burr, and D. Psaltis, "System metric for holographic memory systems," *Optics Letters* **21**(12), pp. 896–898, 1996.