Density implications of shift compensation postprocessing in holographic storage systems

Laure Menetrier and Geoffrey W. Burr

We investigate the effect of data page misregistration, and its subsequent correction in postprocessing, on the storage density of holographic data storage systems. A numerical simulation is used to obtain the bit-error rate as a function of hologram aperture, page misregistration, pixel fill factors, and Gaussian additive intensity noise. Postprocessing of simulated data pages is performed by a nonlinear pixel shift compensation algorithm [Opt. Lett. **26,** 542 (2001)]. The performance of this algorithm is analyzed in the presence of noise by determining the achievable areal density. The impact of inaccurate measurements of page misregistration is also investigated. Results show that the shift-compensation algorithm can provide almost complete immunity to page misregistration, although at some penalty to the baseline areal density offered by a system with zero tolerance to misalignment.

OCIS codes: 210.2860, 100.0100.

1. Introduction

The promise of a next-generation optical storage technology that offers both high density volumetric storage and fast page-oriented data output has driven recent research in holographic data storage.[1] In the past few years, an equivalent areal density of 250 Gbit/sq. in.[2] and an optical readout rate of 10 Gbit/s[3] have been experimentally demonstrated separately, and 150 Gbit/sq. in. and 1 Gbit/s simultaneously.[4]

The success of these experiments has been driven by the ability to extract data from large pixelated data pages (10^6 pixels) despite the presence of both crosstalk and random noise. Crosstalk—between data pixels as well as between data pages—increases as the volume used to holographically record N bits of data shrinks toward the theoretical limit of $N\lambda^3$. Increasing either the readout rate or the number of superimposed holograms reduces signal levels, bringing them closer to the noise floor set by optical scatter noise, Johnson noise in the detector electronics, and shot noise. Tradeoffs between these factors can be optimized by maximizing the overall density at constant bit-error rate (BER). For instance, Bernal *et al*. studied the impact of band-limiting aperture and input/output pixel fill factors by using equivalent areal density to balance the tradeoff between random noise and inter-pixel crosstalk due to diffraction.[5] Similarly, Neifeld and Zhou have studied the impact of the shape of the modulating pixel area.[6]

In practice, however, additional inter-pixel crosstalk—beyond that due to diffraction of the data-bearing wavefront by the finite transverse aperture—will be present. Some of these added crosstalk effects, such as misfocus and most of the optical aberrations (coma, spherical aberration, etc.), act to broaden the point spread function (PSF) beyond the size of the diffraction-limited blur spot. Even more problematic are effects that shift pixel energy away from the intended detector pixel and onto one of its neighbors, such as lateral misalignment, optical distortion, or magnification error. These effects can originate from optical design or fabrication, mechanical motion of optical components, holographic effects such as Bragg-degenerate reconstruction,[7] or media effects such as shrinkage in photopolymers.[8] At best, these additional crosstalk factors merely consume a larger fraction of the signal-to-noise ratio (SNR) budget, requiring an increase in signal levels and a reduction in performance (slower readout or lower density). At worst, these additional factors push the BER in the absence of random noise over acceptable levels, bringing the achievable storage density to zero.

The authors are with IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120. E-mail address for G. W. Burr is burr@almaden.ibm.com.

Received 17 May 2002; revised manuscript received 22 October 2002.

0003-6935/03/050845-16$15.00/0

Recently, a nonlinear signal processing algorithm was introduced to reduce interpixel crosstalk in holographic storage systems.[9] This shift compensation algorithm differs in two ways from most two-dimensional (2-D) postprocessing methods previously developed for holographic data storage.[10–18] First, shift compensation is built around a nonlinear channel model. While a linear channel model would draw upon the extensive linear systems techniques developed for conventional communications and storage channels, the physical detection of coherent light with spatially-integrating intensity detectors is inherently nonlinear.[9,15,19] Second, these earlier signal processing methods for holographic storage, mostly designed to correct pixel blur, fared less well when a significant pixel shift was also present. (One exception to this was the decision feedback–Viterbi algorithm proposed by Heanue *et al.*,[10] which could correct for a shift along one of the two lateral dimensions.) In contrast, the shift-compensation algorithm was expressly designed to correct pixel shifts—specifically, the combination of a known pattern of local pixel shifts (e.g., from optical distortion or material shrinkage) and a dynamically measured global page shift.[9] As part of this process it tends to equalize some of the pixel blur as well. Because misalignments by an integer number of pixels are simple to correct (requiring only bookkeeping in reassigning pixels and redundant encoding of data at opposite edges of the page), an algorithm that could successfully compensate for a residual fractional shift of up to and including ± 0.5 pixel could easily be extended to any arbitrary shift.

When implemented on the DEMON II holographic storage platform,[2] the original shift compensation algorithm relaxed the tolerance to which megapel data pages had to be aligned (to maintain $<10^{-3}$ raw BER) from $\pm 16\%$ to $\pm 40\%$ of the pixel pitch.[9] (Note that a raw BER of 10^{-3} is low enough that error-correction coding of moderate overhead and complexity can still deliver a user BER $<10^{-12}$.[20]) While qualitatively crucial to the 250 Gbit/sq. in. and 150 Gbit/sq. in. + 1 Gbit/s experiments demonstrated on DEMON II,[2,4] the shift-compensation algorithm did not in fact provide complete immunity to page misalignment. When all the contributing factors at any given spot on the data page (global misalignment plus distortion) led to a total local shift of ~ 0.5 pixels, noise amplification and error propagation appeared to push the associated raw BER over 10^{-3}.[9] However, neither the interplay between this interpixel crosstalk and the amount of random noise present, nor the resulting effects on the achievable density were explored in these experiments.

In this paper, we use a numerical simulation to quantify the relationship between data page misalignment, random noise and shift compensation, and to explore the resulting impact on raw BER and achievable areal density. By restricting our study to global misalignment of the entire page and then plotting all results along the two important shift–symmetry axes, we provide enough information to allow our results to be extended to any arbitrary pattern of local shifts due to optical distortion, magnification error, etc. Section 2 contains a description of our numerical simulation, in which fast Fourier transforms (FFT)(s) and oversampling are used to calculate the field amplitude at the detector plane, when imaging or holographically retrieving a small pixelated data page. The resulting raw BER after detection is calculated as a function of hologram aperture, page misregistration, pixel fill factors, and Gaussian additive intensity noise at the detector pixel. In Section 3, we relate the amount of detector noise required to push the raw BER over a given level to the achievable areal density. The joint effects of hologram aperture and page misregistration on storage density are shown. In Section 4, we add shift-compensation postprocessing to the simulation, and show the resulting improvements in BER and storage density. Section 5 describes the effects of operating the shift compensation algorithm with an incorrect estimation of the local shift. Then, Section 6 shows a comparison of the nonlinear shift-compensation algorithm against the Wiener filter (a best-case linear filter in the presence of noise). Last, in Section 7, we discuss avenues for further improvements to the shift-compensation algorithm and conclude in Section 8.

2. Review of the Method

The numerical simulations in this paper model the 4-*F* optical system. Not only can this configuration be simulated in a straightforward manner with the FFT,[5] but 4-*F* systems are also widely used in real holographic storage systems.[1,2,21]

A 4-*F* system consists of two lenses separated by the sum of their focal lengths, with input and output planes separated by twice this distance. Initially, we assume that these focal lengths f are identical, providing unity magnification between input and output devices of identical pixel pitch. (Such one-to-one pixel matching can be achieved for unequal pixel pitches if the ratio of the focal lengths matches the ratio of pixel pitches.) The input spatial light modulator (SLM) contains pixels on a square grid spacing of Γ and linear fill factor of g_{SLM}, making the active modulating area of each SLM pixel $g_{\mathrm{SLM}}\Gamma \times g_{\mathrm{SLM}}\Gamma$. The output detector array has identical pitch and a linear fill factor of g_{CCD}. Although the detector array is identified as a CCD detector, these simulations apply equally well to CMOS or other detector arrays. A square aperture of area D^2 is placed in the common focal plane of the two lenses. We assume that the system has a space-invariant impulse response (PSF) that is due solely to the aperture. In a practical system, there may be more than one PSF, each valid over a large patch of the detector array.[22] Defocus, lens aberrations, and magnification error will not be treated, but lateral misalignment will be introduced below as necessary.

The pixel spread function of this 4-*F* system is the convolution of the PSF (e.g., the Fourier transform of the square aperture) with the original pixel shape.[5]

If the aperture is thought of as a low-pass filter passing spatial frequencies within $\pm D/2$, and the pixel spacing at the CCD camera as a sampling at the frequency $\lambda f/\Gamma$, then the square aperture with side $D_N \equiv \lambda f/\Gamma$ corresponds to the Nyquist sampling condition.[5] With temporal signals this condition is usually met if one chooses to sample at twice the known signal bandwidth, then low-pass filter to this bandwidth. In this context the sampling rate is fixed by the SLM pixel spacing, and it is the low-pass filter of Nyquist aperture D_N that is chosen. Describing aperture sizes in units of D_N allows simulation results to be independent of the particular choice of λ, f, and Γ up to the point at which an absolute areal density is evaluated.[5]

The 4-F optical system can be accurately modeled for a small data page by applying a 2-D FFT, a 2-D low-pass operation, and a second 2-D FFT.[5,15,19] Each input SLM (and CCD) pixel was oversampled by a grid of 40 $\times$ 40 subpixels. To center the active pixel area properly, fill factors were quantized to the nearest even number of subpixels. The largest matrix we could efficiently put through the FFT algorithm was 2048 $\times$ 2048, corresponding to 51 $\times$ 51 pixels with 8 additional blank subpixel rows and columns. To increase the page size available for shift compensation, pages of 186 $\times$ 186 pixels were used, with the central 180 $\times$ 180 pixels populated with random binary data under equal *a priori* probabilities (i.e., roughly the same number of ON and OFF pixels). For the FFT, each page was broken into sixteen subpages containing 45 $\times$ 45 data pixels and a three-pixel wide ring of the appropriate neighboring pixels from the large page. Because no space-variant effects were modeled, each subpage was simply imaged as if it were on axis and then reassembled at the far end. An SLM intensity contrast of c was applied by setting the field amplitude within each subpixel of the OFF pixels to $a_0 = (40 g_{\mathrm{SLM}} \sqrt{c})^{-1}$.

After the first FFT, a band-limiting aperture was applied. Because the input grid contained a grid of 51.2 $\times$ 51.2 pixels, the Nyquist aperture corresponded to the central 51.2 $\times$ 51.2 samples of the 2048 $\times$ 2048 Fourier plane matrix representation. To keep apertures centered properly, each desired aperture was rounded to the nearest odd number of samples. Thus simulations run at the Nyquist aperture D_N actually correspond to 0.9961 $\times$ D_N. However, even at small apertures the resulting effects on the simulation results were overshadowed by the statistical variations due to the random page encoding (e.g., the occurrences of various pixel neighborhoods with good and bad crosstalk properties[23,24] fluctuate from run to run).

As in Ref. 5, we normalized the total output power so that each SLM pixel turned ON delivers a normalized signal of 1.0 through the aperture. Including the SLM intensity contrast c, the power after the aperture was $P = N_1 + N_0/c$, where N_1 and N_0 are the number of ON and OFF pixels, respectively. Thus a low CCD fill factor reduces the detected signal values, but a small aperture or a low SLM fill factor does not. In practice this implies that, after optimizing the modulation depth, a small aperture or a low SLM fill factor must be counteracted by longer hologram exposures to reach the same diffraction efficiency in the end. When the $M/\#$ of the system (the constant of proportionality between the diffraction efficiency and the number of holograms squared[25]) is independent of the aperture (transmission geometry but not 90° geometry[1]) the result is a loss of recording rate rather than of dynamic range and density.

A second FFT produced the optical field amplitude that would arrive at the surface of a properly focused detector array. Each pixel of the virtual CCD camera integrated the resulting intensity (field amplitude squared) over its active area, represented by $(40 g_{\mathrm{CCD}})^2$ subpixels. To include the effects of lateral misregistration, the range of subpixels over which each pixel's signal was accumulated was simply offset. The lateral shift of the detector array in the x and y directions (δ_x and δ_y, respectively) is thus expressed in fractions of the pixel pitch Γ, quantized to the nearest 1/40th of a pixel.

Interestingly, this simulation of a holographic storage system does not explicitly include a hologram. The detected wavefront could either be directly imaged from the SLM, or recorded and reconstructed from a hologram placed somewhere in the optical path. By assuming that the storage and reconstruction of this data-bearing wavefront by a hologram does not significantly alter it, we implicitly neglect any optical aberrations introduced by the holographic media, inter-page crosstalk, optical scatter, further band limiting from an insufficiently large reference beam, the effects of material absorption and recording nonlinearities,[26] and other effects such as beam coupling or photovoltaic-induced index changes.[1,2]

Because we retained the *a priori* information from the original SLM pattern, the signals obtained for each detector pixel could be separated into the two appropriate classes (ON or OFF pixels). Accumulating the occurrences from across the small data page in a finite number of brightness bins created two discrete histograms representing the interpixel crosstalk in the absence of noise. Two example histograms are shown in Fig. 1, for horizontal shifts of 0.25 and 0.5 pixels. For calculations, a typical choice of bin size was 1/1000 of the normalized output power (e.g., the 1.0 from the normalization at the Fourier transform aperture). For better visualization of the distributions, the bins shown in Fig. 1 are 10$\times$ broader. Once a single intensity threshold can distinguish these two classes correctly, then no errors are detected. However, ON and OFF intensity distributions that are far away from each other would seem to be more desirable than those that are close but do not quite overlap. To quantify this intuition, we consider the added impact of random noise in the presence of this interpixel crosstalk.

The random noise considered in this study is zero-mean additive Gaussian intensity noise, corresponding physically to thermal fluctuations (Johnson noise) in the detector electronics. This noise is often the

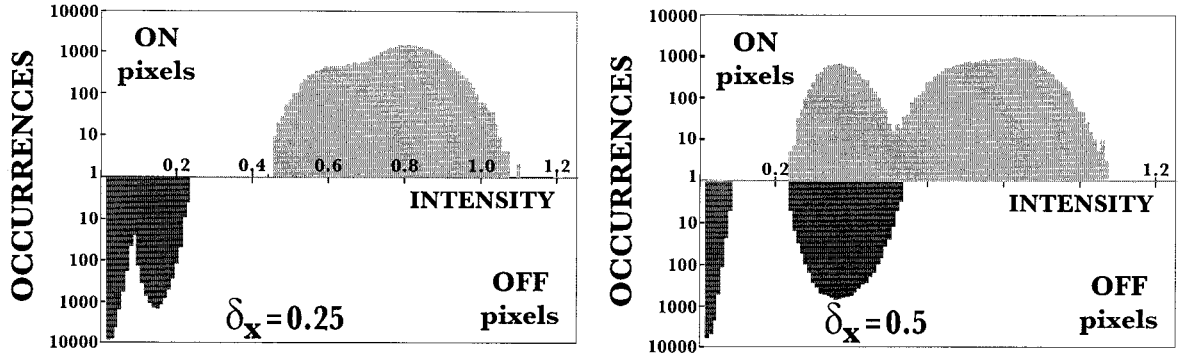

Fig. 1. Histograms for ON and OFF pixels after passing through a 4-*F* system containing a square aperture with sides 1.4× those of the Nyquist aperture, D_N. The detector array is misaligned: (a) by 0.25 pixels, (b) by 0.5 pixels along x. To show the distributions clearly, the occurrences of OFF pixels are plotted along the negative vertical axis.

limiting random noise source in holographic storage systems, especially for photorefractive media, where the level of optical scatter from the holographic media is low.[1] (If both Johnson and scatter noise could be reduced sufficiently, then the underlying shot noise would become the dominant random noise source.) We parameterized the detector noise by its standard deviation, σ_d, given as a percentage of the normalized output power. To model the statistical fluctuation from random noise, we convolved each intensity bin in the discrete histogram with this Gaussian distribution.[5]

We define the BER as the probability that a pixel will be classified incorrectly by a single threshold Θ, written as[5]

$$\mathrm{BER}(\Theta) = \frac{1}{4}\left[\frac{1}{d_0}\sum_{i=1}^{N} w_0(i)\,\mathrm{erfc}\left(\frac{\Theta-\mu_i}{\sqrt{2}\sigma_d}\right) + \frac{1}{d_1}\sum_{i=1}^{N} w_1(i)\,\mathrm{erfc}\left(\frac{\mu_i-\Theta}{\sqrt{2}\sigma_d}\right)\right]. \quad (1)$$

Here N is the total number of bins filled, $w_0(i)$ $[w_1(i)]$ is the number of elements of the OFF (ON) distribution in the ith bin, d_0 and d_1 are the total number of counts in each distribution, μ_i is the intensity corresponding to the ith bin, and erfc is the complementary error function. In a practical system, mistakes made in choosing an optimal threshold without access to *a priori* information inevitably introduce additional error.[21] In practice, however, we have found that the benefits of applying multiple local thresholds (which can differ) roughly balances the loss of *a priori* information, making optimal global thresholding an accurate indicator of BER trends for practical decoding algorithms.[21] Thus the minimum BER is simply evaluated at the optimal value of Θ found by repeated evaluation of Eq. (1).

Figure 2 shows the calculated BER as a function of the lateral shift along the x and $y = x$ axes for several different apertures between 1× and 2× the Nyquist aperture. Note that because of symmetry, the left-hand portion of Fig. 2 shows the BER that would be expected for shifts along any of four axes: the $\pm x$ axes and the $\pm y$ axes (see figure inset). Similarly, the right-hand side panel shows the somewhat higher BER to be expected for any of the four diagonal shifts such that $|\delta_x| = |\delta_y|$. (To see the difference between the two subplots clearly, compare the intersection points for each curve with the BER = 10^{-3} line). Thus the BER at any arbitrary shift (δ_x, δ_y) within ± 0.5 pixels can be reasonably interpolated from—and certainly bounded by—the BER values on the nearest diagonal and vertical/horizontal axes, for which the data are compactly given by Fig. 2. Similar plots will be used throughout the remainder of the paper, providing the opportunity for the reader to estimate the overall performance for a global shift that is accompanied by any given spread of local shifts from distortion. Had we ourselves picked some arbitrary distortion pattern to study, such extraction to any other distortion pattern of interest would not be possible.

Other parameters for Fig. 2 are set to values typically seen in experimental demonstrations ($g_{\mathrm{SLM}} = g_{\mathrm{CCD}} = 0.9$, $\sigma_d = 0.05$, $c = 100$). Note that as would be expected, the BER tends to increase with added interpixel crosstalk, whether due to smaller apertures (a broader PSF) or lateral misalignment.

A BER value of particular interest, marked with a dashed line, is at 10^{-3}. This is the raw BER at which the overhead associated with error-correction coding begins to grow from moderate to excessive. For instance, a Reed–Solomon code with 255 eight-bit symbols per codeword can deliver 10^{-12} user BER from a raw BER of 10^{-3} with an redundancy overhead of only 10%[20] and manageable implementation complexity (e.g., commercial chips are available). However, a raw BER of 10^{-2} requires a redundancy overhead of nearly 40% and is significantly more difficult to implement. Simultaneous optimization of the error correction and shift compensation to optimize coding gain is beyond the scope of this paper. While this might lead to a slightly different target raw BER, we expect that the overall density trends would remain the same.

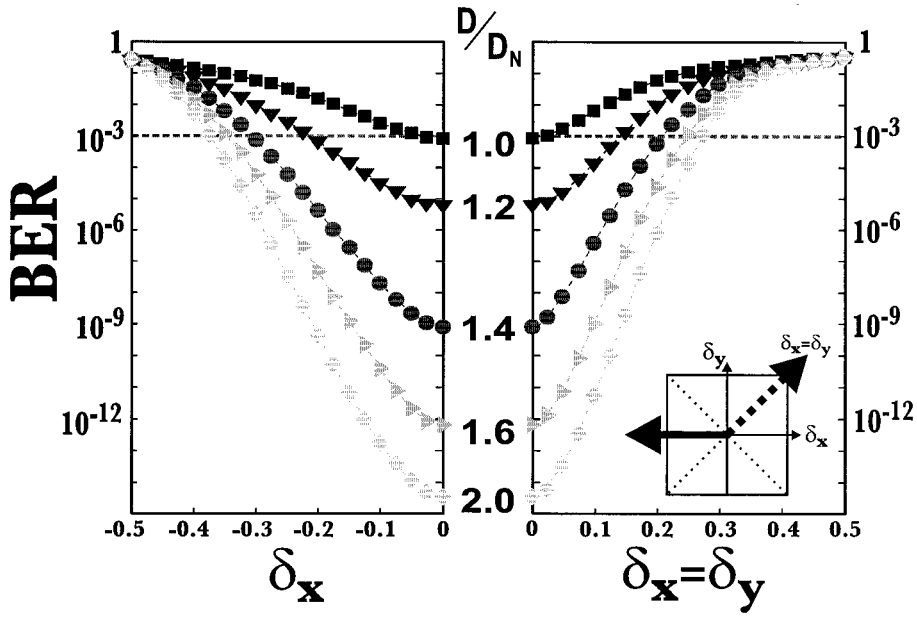

Fig. 2. Simulated bit-error rate (BER) as a function of the page shift: (a) along the x axis ($\delta_y = 0$), (b) along the line $y = x$ ($\delta_y = \delta_x$), for various bandlimiting apertures. Page shifts are in units of a pixel, aperture sizes in units of the Nyquist aperture. The device fill factors are large ($g_{\mathrm{SLM}} = g_{\mathrm{CCD}} = 0.9$), SLM contrast is high ($c = 100$) and a moderate amount of detector noise is present ($\sigma_d = 0.05$).

Figure 2 shows that, given this level of random noise, a holographic storage system with a small aperture ($D \sim D_N$) must be precisely aligned to achieve this target BER level. For larger apertures, a system need only maintain the page alignment within a given circle of tolerance around perfect registration to keep the BER at or below 10^{-3}. While a larger aperture increases this tolerance to misalignment, a shift of half a pixel always leads to excessively high BER.

3. Density versus Misalignment

Given this procedure for evaluating BER, we can now solve for the achievable areal density. The density is simply the amount of data stored in a stack of superimposed holograms divided by the aperture area D:

$$\mathcal{D} = \frac{N_p M}{D^2}, \tag{2}$$

where M is the number of stored holograms and N_p the number of pixels per hologram.

For any given D (and N_p), storage density can be maximized by increasing the number of holograms M until the BER approaches 10^{-3}. Because the diffraction efficiency of M holograms scales as $\eta = [(M/\#)/M]^2$, an increase in the number of holograms implies a decrease in the diffracted power, and thus a reduction in the number of signal electrons detected at each detector pixel, n_s. Given that the equivalent number of detector noise electrons, n_d, remains constant, the ratio between signal and noise electrons has just decreased. We can relate this to the Gaussian noise variance introduced in the previous section by $n_d = \sigma_d n_s$. Combining these equations with the dependence of signal electrons on diffraction efficiency ($n_s \propto \eta$), we can quickly write that

$$M \propto \sqrt{\sigma_d} \qquad \mathcal{D} \propto \frac{\sqrt{\sigma_d}}{D^2}. \tag{3}$$

Thus a holographic storage system that can tolerate four times as much random noise (with the same aperture D) can store twice as many holograms and achieve twice as much areal density.

In the numerical simulations, we evaluated the areal density by increasing the amount of random noise (σ_d) from zero until the BER reached 10^{-3}. The resulting areal density, in arbitrary units, is then simply obtained from Eq. (3). If the BER exceeds 10^{-3} before any random noise is added, then the achievable areal density is considered to be zero. Figure 3 shows a plot of the resulting density as a function of the page shift for various apertures between $1\times$ and $2\times$ the Nyquist aperture. As before, a single plot is used to illustrate the impact of shifts along both the x and $x = y$ axes. These simulations reiterate the conclusions of Ref. 5: In a perfectly aligned system, areal density is maximized at an aperture just larger than the Nyquist aperture.

However, there is an interesting tradeoff between the achievable density and the tolerance to lateral misregistration. The best densities are achieved at small apertures, but this is exactly where the system is most sensitive to misalignment. Tolerance to misalignment requires sacrifices in areal density, but even a large aperture (twice the Nyquist aperture) delivers no density at a half-pixel shift. And misregistration in both x and y are commensurately more difficult to correct. Each aperture size has a range of misalignments for which it is the optimal choice, as

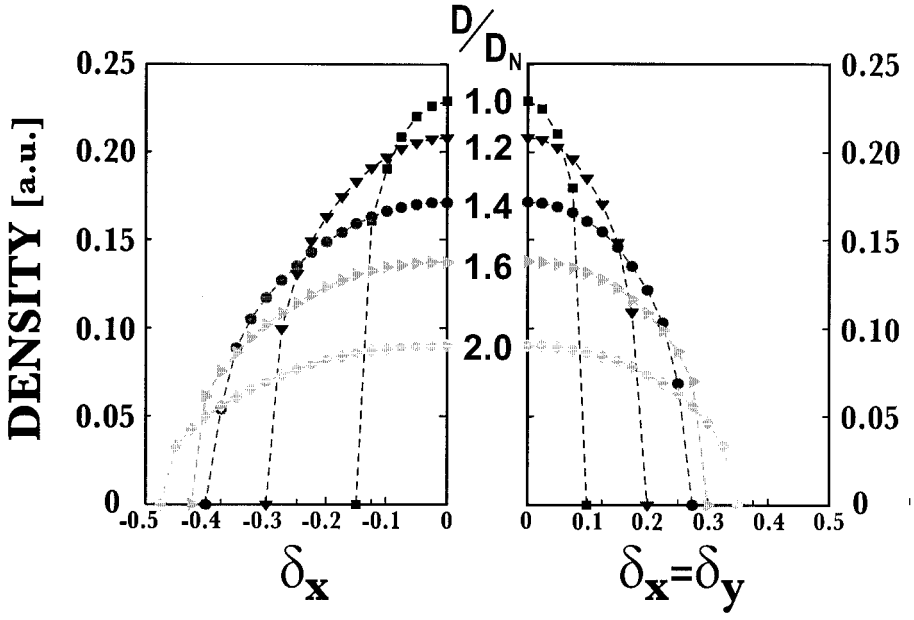

Fig. 3. Simulated areal density in arbitrary units as a function of the page shift: (a) along the x axis ($\delta_y = 0$), (b) along the line $y = x$ ($\delta_y = \delta_x$), for various bandlimiting apertures. Page shifts are in units of a pixel, aperture sizes in units of the Nyquist aperture. The device fill factors are large ($g_{\mathrm{SLM}} = g_{\mathrm{CCD}} = 0.9$), SLM contrast is high ($c = 100$). Density is obtained by increasing random noise until a BER of 10^{-3} is achieved, zero density implies that interpixel crosstalk has already driven the BER over this target even without any random noise.

the portion of the SNR budget given over to combat interpixel crosstalk is shifted from blur to misregistration.

4. Shift Compensation

Figure 3 quantifies the reduced tolerance to misalignment that accompanies high areal density, an effect that we had qualitatively observed while working with the DEMON II experimental holographic storage platform.[2] This platform used custom optics to image a 1024×1024 page through an aperture of $1.36 \times D_N$. Because the wavelength λ was 532 nm, the focal length $f = 30$ mm, and the pixel pitch Γ_{SLM} was 12.8 μm, the Nyquist aperture D_N had sides with a length of 1.25 mm. Storing 1000 holograms at each of several neighboring locations, spaced on a grid of 1.6 mm over the entrance face of the $LiNbO_3$:Fe crystal, eventually produced an experimentally demonstrated (channel) areal density of 394 pixels/μm^2 (254 Gpixels/sq. in.).[2]

Initially, however, it was difficult to align the system well enough to image the megapel page through the customized 4-F system, let alone reconstruct low-error holograms. Despite the careful optical design and fabrication, it turned out that residual optical distortions in the corners of the data page could reach as high as 0.25 pixels. Fortunately, the width of the PSF varied little even as the center position of each pixel's image strayed significantly from the desired square grid. But even under the best-case alignment, some portion of the page was always locally misaligned, leading to decoding errors and an increased page-wide BER. This motivated us to turn to the signal-processing literature that was being produced for holographic storage.

These studies offered an impressive variety of algorithms and techniques,[10–15] which had two things in common: They all concentrated primarily on correcting blur rather than pixel shift, and they all were implemented with a linear channel model. It was relatively clear that, given a sufficient support constraint (size of the convolution kernel), most of these techniques could be extended to shift rather than blur. However, a simple experiment made it clear that the linear channel model was inappropriate, while also offering an explanation for why these techniques were providing smaller performance improvements in holographic storage systems than had been expected, given their great success in conventional communications and data storage.

The simulation described above can be used to repeat this experiment. The impulse response and its superposition properties are gauged by sequentially imaging three mostly-OFF data pages through the system. In the first page, only a single pixel is turned ON. In the second, only the immediate (say, horizontal) neighbor of this pixel is turned ON, and in the third, both are ON. The results (the five nearest detector pixels along that row) are shown in Table 1, for an aperture of $1.4 \times D_N$ and a pixel shift of $\delta_x = 0.5$ (CCD and SLM fill factors 90%, infinite contrast, no random noise). If this imaging system were linear, then superposition would hold and the signal detected with both pixels ON would be completely predicted by the first two measurements. The top

Table 1. Simulated Signal Values[a]

Input		Detector Pixels				
. . . 0100 . . .	A	7	364	364	7	1
. . . 0010 . . .	B	1	7	364	364	7
. . . 0110 . . .	C	6	294	939	294	6
		After shift compensation				
	A	10	699	10	2	0
	B	2	10	699	10	2
	C	42	751	770	45	6

[a]Simulated signal values along one row of the detector array under an x misalignment of 0.5 pixels ($D = 1.4\ D_N$, infinite contrast, 90% fill factors, no random noise, normalized signal corresponds to 1000). Row A shows the result for a single ON pixel, row B for the immediate horizontal neighbor of this pixel turned ON, and row C for both pixels ON. The added signal in the center detector pixel of row C is due to the interference term between the overlapped optical field amplitudes, demonstrating the need for a nonlinear channel model. The results of postprocessing each of these rows with the shift-compensation algorithm are also shown.

part of Table 1 shows that—in terms of the detector signal—the system is not linear.

The explanation for this is rather simple: the light is coherent and the system is linear in terms of optical field amplitude, but each detector pixel integrates intensity over its active area. When the detector is aligned, then most of the signal comes from the intended SLM pixel. The incident optical field is the original rectangular SLM pixel convolved with the PSF of the optical system. With a square aperture, the PSF is a sinc function, and the resulting pixel-spread function varies along the x axis as

$$h(x) \equiv c \int_{-g_{\mathrm{SLM}}/2}^{g_{\mathrm{SLM}}/2} \mathrm{sinc}\left[\frac{D}{D_N}(x - x')\right] \mathrm{d}x', \qquad (4)$$

where $\mathrm{sinc}(x) \equiv \sin(\pi x)/(\pi x)$. Here, c is chosen such that $\int_{-\infty}^{\infty} h^2(x)\mathrm{d}x = 1$, x and x' are in units of the pixel pitch Γ, g_{SLM} is still the linear SLM pixel fill factor, and D and D_N the widths of the Fourier-transform aperture and Nyquist aperture, respectively. Note that the integration over the y dimension is omitted only for simplicity, and both the real system and the simulation deal with true 2-D pixels at all times.

As the detector array shifts in a particular direction (say, along negative x relative to the SLM), then most of the crosstalk originates from just one of the neighboring SLM pixels (the one that the detector pixel is moving toward). Referring to Fig. 4, the signal received at a detector pixel r_2 due to both the correct SLM pixel p_2 and its neighbor p_1 can be written as

$$r_2 = \int_{-g_{\mathrm{CCD}}/2}^{g_{\mathrm{CCD}}/2} [\sqrt{p_2}h(x - \delta_x) + \sqrt{p_1}h(x - \delta_x + 1)]^2 \mathrm{d}x. \qquad (5)$$

Here, g_{CCD} is the fill factor of the detector pixel, and δ_x is the total local offset between the SLM image and detector due to misregistration, optical distortion, magnification error, and material shrinkage[8] combined. We will refer to the shift shown in Fig. 4 as a positive δ_x, as though the SLM had shifted along the positive x axis with respect to a stationary detector array. If the shift had been in the other direction, then δ_x would be negative, the $+1$ offset in Eq. (5) would change sign, and p_3 would be the relevant neighbor pixel appearing in the place of p_1 in Eq. (5).

The inherent assumption—that only one neighbor contributes crosstalk—becomes less accurate as the width of the PSF approaches the pixel pitch (e.g., as the aperture size is reduced toward the Nyquist ap-

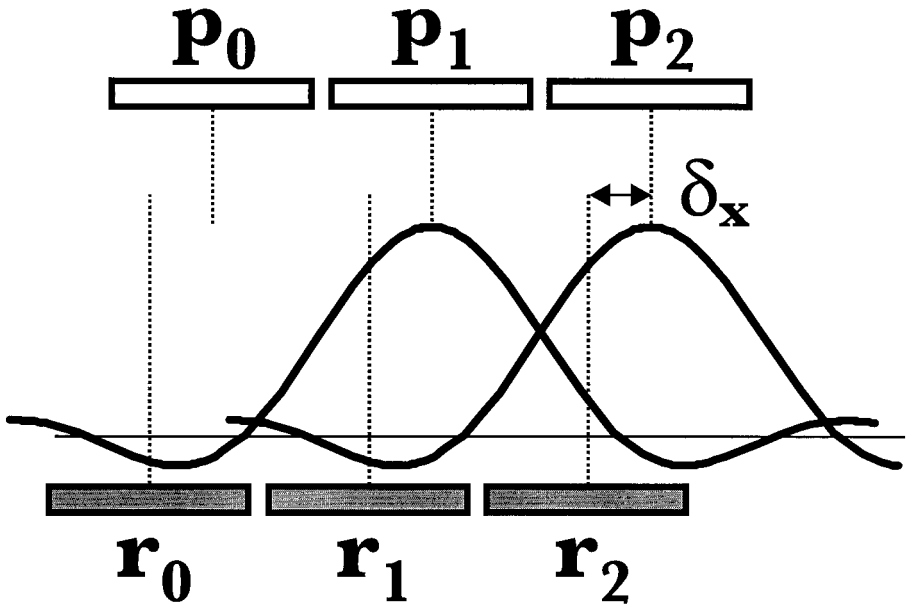

Fig. 4. Spatially blurred images of three SLM pixels (p_0, p_1, p_2) are slightly shifted relative to three CCD pixels (r_0, r_1, r_2), reducing the signal in the desired target pixels and creating crosstalk in their neighbors.

erture), and this can become a problem that leads to poorer performance at small apertures. The one-neighbor assumption is also not appropriate for small shifts, but fortunately very little crosstalk is introduced at small shifts (when the aperture is large), so this is generally not a problem.

Interestingly, if the detector fill factor were to shrink to zero, then the integrand of Eq. (5) would simply be evaluated at the center of the detector pixel. Then under those conditions, one could take the square root and return to a system that was linear in magnitude. Because phase information is still lost, this approach requires that the phase does not vary from pixel to pixel, which is often the case unless a random phase mask has been intentionally introduced. This magnitude-linear approach was tested by Vadde *et al.*,[15] and as expected, it worked well at small fill factors but poorly at large ones. Unfortunately, the need for raw signal strength to overcome the random noise floor strongly favors large fill factors, despite their less favorable interpixel crosstalk properties.[5]

By rewriting Eq. (5) as

$$r_2 = p_2 H_{00}(\delta_x) + 2\sqrt{p_1 p_2} H_{01}(\delta_x) + p_1 H_{11}(\delta_x), \quad (6)$$

the detected signal is decomposed into linear contributions from the two pixel intensities and a nonlinear factor through their constructive interference. The weights $H_{00}(\delta_x)$, $H_{11}(\delta_x)$ and $H_{01}(\delta_x)$ 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, as a function of the known local shift δ_x. The corresponding destructive interference demanded by conservation of energy is distributed across the next ring of pixels and is ignored as a second-order effect. Some of this destructive interference can be observed in the third row of Table 1, in the decreased values measured by the outer detector pixels.

Equation (6) represents a condensed representation for the corrupted output r_2 that is produced by the combination of some arbitrary pixel pattern (p_1 and p_2) and a known shift δ_x. Reference 9 introduced the idea of a shift-compensation algorithm, whose objective is to invert this equation, by using the measured detector signal r_2 and the known shift δ_x to compute the original value of p_2. The values of the three H_{xx} terms can be obtained either by calculation or by measurement. For instance, the value H_{00} could be computed from

$$H_{00}(\delta_x) \equiv \int_{-g_{CCD}/2}^{g_{CCD}/2} [h(x - \delta_x)]^2 \mathrm{d}x. \quad (7)$$

Alternatively the three H_{xx} terms can be measured as a function of δ_x in the actual experimental apparatus. For instance, the numerical simulation introduced in the previous sections can measure these coefficients in the same way Table 1 was produced: For each shift δ_x, a data page with a single pixel is passed through the optical system and detected. The signal received at the desired detector pixel is H_{00}, while that received at the neighboring detector pixel is H_{11}. Again, the direction of the shift dictates which of the two neighbors is relevant. H_{01}, the nonlinear term, cannot be obtained directly. However, when two neighboring SLM pixels are input and the detector is shifted by a nonzero σ, the interference pattern between these two SLM pixels is present in the signal r_{cal} measured by the detector pixel sitting between their arriving image points. Then, given that $H_{00}(\delta_x)$ and $H_{11}(\delta_x)$ are known, $H_{01}(\delta_x)$ is simply

$$H_{01}(\delta_x) = [r_{cal} - H_{00}(\delta_x) - H_{11}(\delta_x)]/2. \quad (8)$$

Once the H_{xx} values are obtained and the appropriate value of r_2 measured, the only unknowns remaining in Eq. (6) are p_2 and p_1. If p_1 is known, then Eq. (6) can be used to solve for p_2, by writing it as a second-order polynomial in $\sqrt{p_2}$ subject to the binomial equation, producing

$$\sqrt{p_2} = \frac{1}{H_{00}(\delta_x)} [-\sqrt{p_1}\, H_{01}(\delta_x) + (p_1\{[H_{01}(\delta_x)]^2 - [H_{00}(\delta_x) H_{11}(\delta_x)]\} + [H_{00}(\delta_x) r_2])^{1/2}]. \quad (9)$$

To obtain the value of p_1 to insert into this, a version of Eq. 9 can be written to solve for p_1 in terms of r_1 and p_0. But at some point, the algorithm must be seeded by prior knowledge about one of the pixels. The easiest way to do this is to leave a blank pixel at the end of the row. Thus, if pixel p_0 has been intentionally left OFF, then p_1 is simply $r_1/H_{00}(\delta_x)$. This value for p_1 can then be substituted into Eq. (9) to solve for p_2. The shift-compensation algorithm described in Ref. 9 simply proceeds in this way along the entire row, solving for each pixel's true value by using the just-processed neighbor pixel. The bottom half of Table 1 shows the results of this shift-compensation algorithm for each of the three crosstalk signals from the top half of the table. At each pixel, we take the measured signal r, subtract the portion that belonged to the previous pixel, subtract a further portion due to interference, and then factor in the missing signal that should have been here, but which actually fell into the next pixel. While the algorithm requires *a priori* information about the local shift δ_x and the associated weighting factors $H_{xx}(\delta_x)$, no information about the binary pixel states or any special encoding arrangement is required.

Note that a pixel shift of exactly one pixel requires only bookkeeping (r_1 is really p_0, r_2 is p_1, etc.) and blank columns are placed at the page boundaries to prevent data from being lost off the edge of the page and to seed the algorithm as discussed above. By extension, the shift compensation algorithm only needs to work for all shifts in the range $-0.5 < |\delta_x| \leq 0.5$ to achieve complete alignment insensitivity. Unfortunately, when $|\delta_x| > 0.5$, then the missing signal ascribed to the next pixel [the term $p_1 H_{11}(\delta_x)$ from Eq. (6)] becomes large and the algorithm be-

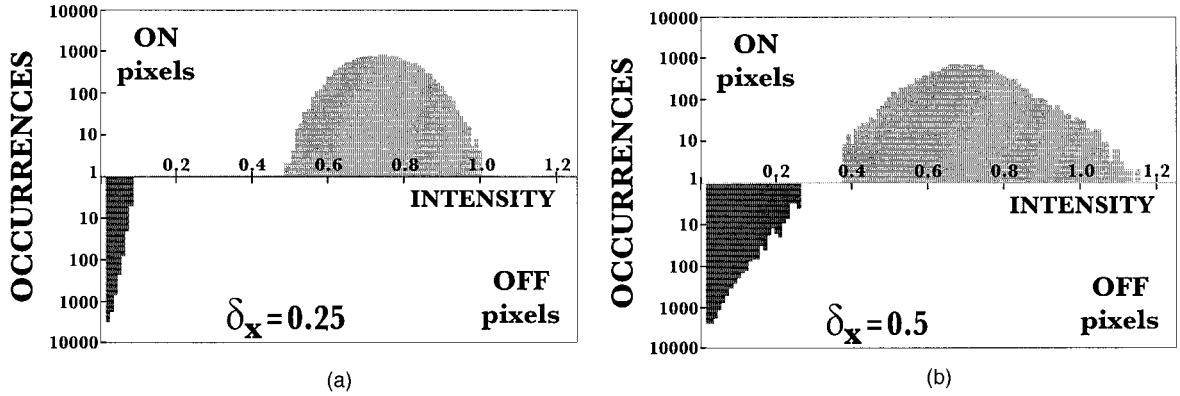

Fig. 5. Histograms for ON and OFF pixels for the same data page as Fig. 1 after shift-compensation postprocessing.

comes prone to error propagation. In essence, because the correction factors to be applied to each pixel become large, any random noise (from scatter or detector noise) at a given pixel is amplified along with the signal to affect the next pixel, leading to more and more error as the algorithm proceeds along the row. Perturbations that can start error propagation include random noise on the pixel values, mistakes in estimating δ_x, or even second-order effects, such as the destructive interference from other neighboring pixels.

For this reason, Ref. 9 described a method of processing each line twice: once from left-to-right to compensate rightward shifts of up to half a pixel (as in Figure 4); and once from right-to-left for leftward shifts. Each pixel is processed on both passes: On one pass, the shift δ_x will have a magnitude <0.5; on the other, by incurring a bookkeeping charge of one pixel the same shift appears as $\mathrm{sgn}(\delta_x)|1 - \delta_x|$ with a magnitude >0.5, where $\mathrm{sgn}(x)$ is the sign of x. This produces good results for shifts of, say, $\delta_x \sim 0.75$, but significant error propagation still takes place at $|\delta_x| \sim 0.5$. To reduce these effects, we soften the transition for any pixel with $|\delta_x| \geq 0.3$ by retaining a weighted average of the estimates from the left-to-right and right-to-left passes. We also performed weighted averaging for $|\delta_x| \leq 0.2$, although the effects here are only observable for small apertures.

This discussion also illustrates why only one of the pixel's two horizontal neighbors is considered. If the effects of both p_1 and p_3 were to be considered in the determination of p_2, then the algorithm would become noncausal. In essence, these estimates of p_1 and p_3 we would need in turn require knowledge of the true value of p_2, making it impossible to perform such a nonlinear algorithm in a single pass. Instead, the algorithm would have to be iterative, successively refining its estimates of pixel values over numerous passes. In fact, such iterative algorithms that can handle nonlinear channels have recently been introduced for holographic storage.[11–13,17,18]

To extend the shift-compensation algorithm to 2-D, one could expand the analysis of Eq. 4 to three neighboring pixels (horizontal, vertical, and diagonal), producing an equation with four linear and six nonlinear terms:

$$\begin{aligned} r_s = {} & p_s H_{ss} + p_h H_{hh} + p_v H_{vv} + p_d H_{dd} + 2\sqrt{p_s p_h} H_{sh} \\ & + 2\sqrt{p_s p_v} H_{sv} + 2\sqrt{p_s p_d} H_{sd} + 2\sqrt{p_h p_v} H_{hv} \\ & + 2\sqrt{p_h p_d} H_{hd} + 2\sqrt{p_v p_d} H_{vd}. \end{aligned} \tag{10}$$

The four pixels in question are termed p_s for self, p_h for horizontal, p_v for vertical, and p_d for diagonal. The approach follows that taken above: the signal received at r_s is a sum of the linear contributions of the three neighbors, plus interference terms associated with all of the pairs of amplitude contributions. Each H_{xx} variable is now a function of both x and y offsets, as in $H_{ss}(\delta_x, \delta_y)$. When both δ_x and δ_y are positive, p_d sits above and to the left of p_s and left-to-right/top-to-down processing produces estimates of p_d, p_v, and p_h before pixel p_s is considered. For each of the three other scenarios, the page must correspondingly be processed in the correct direction. As with the one-dimensional (1-D) variant of the algorithm, each of these H_{xx} parameters can be calculated or measured.

This 2-D variant of the shift-compensation algorithm must be used if aberrations dominate the PSF. However, if the dominant influence on the PSF is diffraction due to an aperture described by functions separable in x and y (e.g., a square or rectangular aperture), then the PSF of the optical system is separable. Under these conditions the shift compensation algorithm can be extended to 2-D by processing each dimension of the data page separately. This is how shift-compensation was implemented on both the DEMON II[2,9] and DEMON III[27] experimental test platforms. Each 2-D page is processed with the 1-D algorithm four times: twice horizontally as described above on all rows to produce an interim page (corrected for all x shifts), and then twice vertically (top-to-bottom and bottom-to-top) on all the columns of this interim page. By postprocessing the simulated 2-D pages that produced the histograms in Fig. 1, we obtained the qualitatively improved histograms shown in Fig. 5. It should be noted that these his-

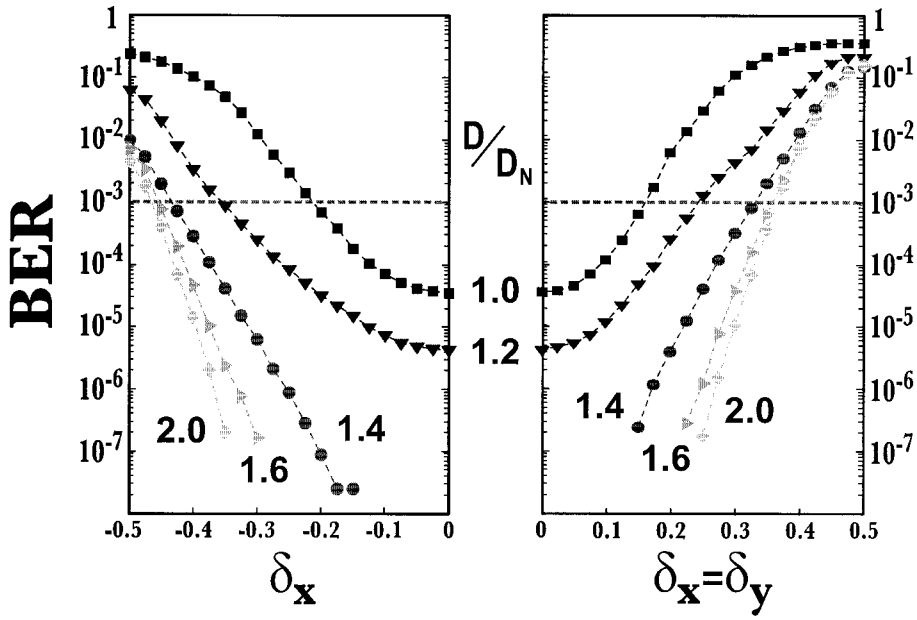

Fig. 6. Simulated BER after shift compensation as a function of page shift: (a) along the x axis ($\delta_y = 0$), (b) along the line $y = x$ ($\delta_y = \delta_x$), for various bandlimiting apertures. System parameters are identical to Fig. 2, enabling a direct before-and-after comparison.

tograms are obtained in the absence of random noise, so the existence of brightness variations after postprocessing means that the shift-compensation algorithm—as it currently stands—does not perfectly invert the data channel. Almost certainly this is due to the inaccuracy of the one-neighbor assumption.

In the experimental implementations, the total lateral shift fed to the shift-compensation algorithm was the sum of dynamically measured global misregistration and a memorized local offset due to magnification error, material shrinkage, and optical distortion.[9,27,28] In these simulations, we restrict ourselves to pure global misregistration, so that the local shift at each pixel is identical to the global page shift. As described earlier, we chose to do this so that the BER results obtained here can straightforwardly be extended to data pages with any arbitrary distribution of local shifts. The expected aggregate BER for a page with varying local shifts can be estimated as the weighted average of the BER for each local shift present, with the weight being the fraction of the page that shares that same local shift. This is a rough estimate, because in practice there will be error propagation between regions of different local shift. However, if we had selected a particular spread of local shifts to study, then our results would provide almost no information about the system performance when broader or narrower distributions of local shifts are present.

To quantify the improvement in system performance, we again wish to add noise and compute the BER. However, it is no longer possible to sum complementary error functions to compute the BER, because it is necessary to perform the shift compensation postprocessing on a noise-perturbed data page and calculate the BER afterwards. Thus we turn to Monte Carlo methods to add and analyze the effects of noise.

Using a random number generator to select values within the appropriate Gaussian probability density function, we perturbed the signal measured at each detector pixel with noise. The resulting output page was postprocessed with the shift compensation, and data values accumulated in running ON and OFF histograms. After a sufficient number of repetitions, the BER was calculated by dividing the number of errors at the best-case intensity threshold by the total number of samples in the histograms. Because the accuracy of this method improves with the number of samples, five different data pages of 180×180 pixels each were used, each subject to between 10 and 500 multiple repetitions of random noise values (e.g., 32,400 random noise samples per repetition). Thus for error-prone pages, the simulation continued until it counted at least ~5000 errors; for low-error pages ($<10^{-4}$), a single bit error corresponded to a BER floor of approximately 1.2×10^{-9}. With this number of repetitions, the run-to-run variation in BER at the desired target level of 10^{-3} was roughly 2%. We compared the results of estimating the BER in the absence of shift compensation with this Monte Carlo procedure against the results of the earlier calculations (using sums of complementary error functions), and found excellent correspondence for all BER values down to 10^{-6}.

Figure 6 shows the improvement in BER after shift compensation over the performance shown earlier in Fig. 2. In general, shift compensation reduces the BER at all shifts, with the worst-case BER still produced at half-pixel shifts. At an aperture of $D = 2D_N$, a BER $<10^{-3}$ can be maintained for all shifts up to nearly $\delta_x = 0.475$, although large shifts in both x

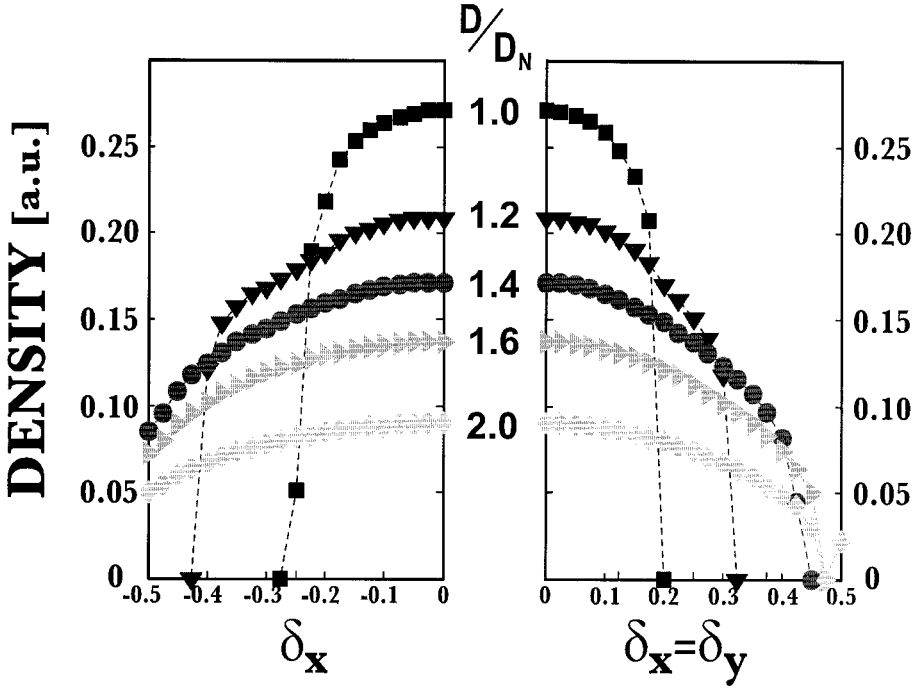

Fig. 7. Simulated areal density in arbitrary units as a function of the page shift: (a) along the x axis ($\delta_y = 0$), (b) along the line $y = x$ ($\delta_y = \delta_x$), for various bandlimiting apertures. System parameters are identical to Fig. 3, enabling a direct before-and-after comparison.

and y still result in numerous errors for this level of noise ($\sigma_d = 0.05$). For small apertures ($D = D_N$), the BER is lower after shift compensation even at zero shift, because the weighted averaging of the two passes helps correct some of the pixel blur.

The achievable areal density can be computed in the same way as before: by determining the level of random noise σ_d required to push the BER over 10^{-3}. However, as we successively approach the correct amount of added thermal noise, the exact crossover point can end up being masked by the run-to-run variations in BER. (These are variations due to randomness in both the pixelated pages and the added noise.) Thus to accurately pinpoint the 10^{-3} crossover point, we used a linear fit through at least six different values of σ_d, covering the BER range from 5×10^{-4} to 2×10^{-3}.

The resulting areal density is shown in Fig. 7. At each aperture, the range of page shifts that can be tolerated has increased (over the range supported without shift compensation shown in Fig. 3), followed by a rapid decrease of the density to zero at some maximum tolerable shift. For apertures of $1.4 \times D_N$ and larger, any shift along a single dimension can be compensated, at a cost in density of roughly 50%. However, large shifts in both x and y still lead to zero density: the high BER due to interpixel crosstalk already exceeds the target threshold of 10^{-3}, with zero tolerance for additional random noise. Thus the required signal levels would be so high (to keep the ratio of signal to random noise very large) that at best only a few holograms could be multiplexed.

However, at an aperture of $2.0\ D_N$, Figure 7 shows that the shift compensation algorithm can provide nearly complete coverage of all possible shifts, with the only gap being at a pixel shift of $|\delta_x| = |\delta_y| = 0.475$. The nonzero density at $|\delta_x| = |\delta_y| = 0.5$ comes about simply because the BER due to interpixel crosstalk at this point is just slightly below 10^{-3}, while at $|\delta_x| = |\delta_y| = 0.475$ is just slightly above 10^{-3}.

Thus by moving the target BER even a little higher (by switching to a slightly stronger error-correction code) or by using a slightly larger aperture, one could achieve acceptable shift-compensation for all possible pixel shifts, albeit at a large cost in density. From Fig. 7, the achievable density would be reduced to 25% of the density achievable in a system with no tolerance to shift and the same aperture ($2D_N$), or less than 10% of the density offered by a shift-intolerant system with an aggressively small aperture (D_N).

However, it should be noted that in some cases, the presence of shift compensation may open up a tradeoff elsewhere in the system that tends to increase density. For instance, one method used to reduce the effects of shrinkage in photopolymer media is to preexpose it before recording, which unfortunately sacrifices some of the potential dynamic range ($M/\#$) of the media. If the increased tolerance to misalignment and local pixel distortions offered by shift compensation can allow significantly more of the inherent dynamic range to be used, then the overall density can actually increase (lower density due to more susceptibility to random noise, but higher due to better $M/\#$). The caveat is that the pattern of pixel distortions caused by this media shrinkage must be known, either by precalibration of determin-

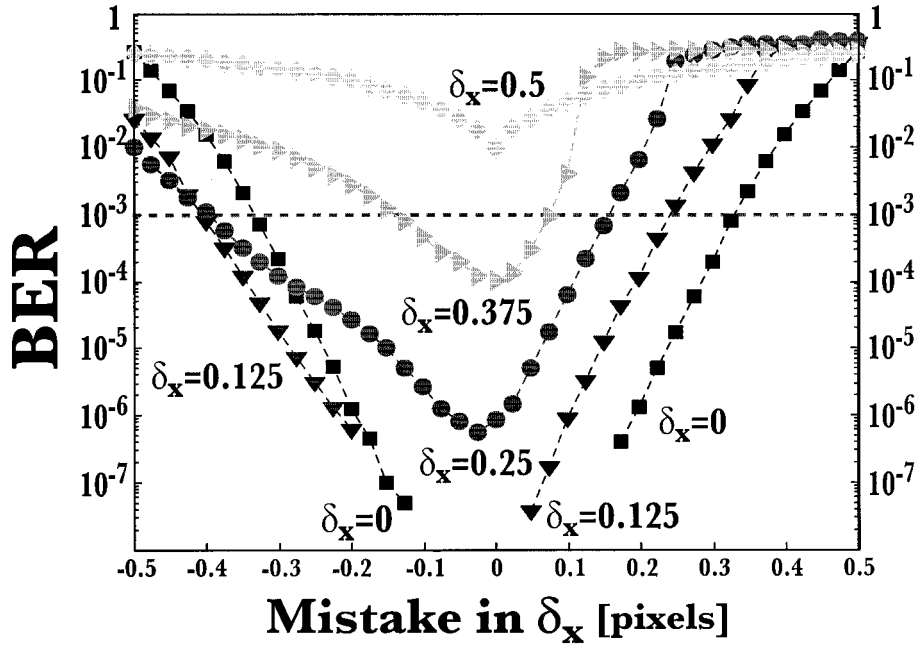

Fig. 8. BER after shift compensation as a function of the error in estimating the pixel shift, for various actual page shifts. System parameters are identical to Fig. 2.

istic changes or by measurement with test patterns recorded along with the actual modulated data.

As a check on our density analysis, we can attempt to convert from the arbitrary units of Eq. (3) to real density units. Reference 5 showed that for a system with half the pixels ON, the achievable areal density could be written as

$$\mathcal{D} = \frac{M/\#}{(D\,D_N)^2}\left[2N_p\,\frac{P_{\mathrm{ref}}}{h\nu}\,\frac{\sigma_d}{(n_d+n_f)}\,\eta_{\mathrm{opt}}\eta_e t_{\mathrm{int}}\right]^{1/2}. \quad (11)$$

The terms σ_d and D as in Eq. (3), as well as the terms $M/\#, N_p$, and n_d remain as defined in Section 3 above. Additionally, $P_{\mathrm{ref}}/h\nu$ is the number of photons per second in the reference beam, t_{int} is the integration time, and η_{opt} and η_e are the efficiency of the optical system and the camera quantum efficiency, respectively.

Now it is possible to forge a connection between Fig. 7 and a real-world experimental demonstration, such as the high-density holographic storage experiment performed by the IBM team on the DEMON II platform.[2] In this experiment, an effective areal density of 394 channel bits/μm^2 was achieved while maintaining a raw BER of 10^{-3}. To reflect the added noise sources in the real experiment, we have introduced an additional noise term in Eq. (11); n_f is the equivalent noise at each pixel due to scatter, inter-page crosstalk, holographic distortions, such as photovoltaic noise, and additional band-limiting by the reference beam. (Note that the correct way to model these factors would be to include them in the simulated data page before detection, and then measure the now smaller value of σ_d at which the raw BER crosses the maximum tolerable limit, but to do this treatment correctly would go well beyond the scope of this work.)

The operating parameters of the DEMON II platform were $M/\# = 1.1$, $P_{\mathrm{ref}} = 345$ mW, $N_p = 10^6$, $\lambda = 532$ nm, roughly $n_d = 100$ noise electrons per pixel, $\eta_{\mathrm{opt}} \times \eta_e = 0.2$, integration time 23.8 ms, and a Nyquist aperture of $D_N = 1.25$ mm. We estimate that because of the significant optical distortion on the data page, the curve of density versus global shift was sharply narrowed, so that the effective operating point on the vertical scales of Figs. 3 and 7 corresponded to $\sim$0.1. Given these parameters, Eq. (11) produces a density of $\sim$ 400 bits/μm^2 when $n_f = 180$, which corresponds well to our qualitative observations of the importance of thermal noise relative to the other noise sources in the DEMON II experiment. Thus we have shown that the relative densities calculated under the assumptions made in this paper can—with care—be related to experimentally demonstrated effective areal densities in units of bits or pixels per μm^2.

5. Compensating the Wrong Shift

One of the assumptions we have made to this point has been that the values of δ_x and δ_y used within the shift compensation algorithm (for instance, to determine the set of H_{xx} values and weighting factors) have been identical to the actual shift that existed at the detector array. However, this implies that the measurement of the x and y page shifts, by the use of fiducial patterns or other means, will always arrive at the correct result. This is an unlikely scenario, especially when fiducial patterns are being measured in the presence of noise.

In Fig. 8, we plot the BER after shift compensation

as a function of the error between the actual shift acting on the pixelated data page and the measured shift used to select the parameters and operation of the shift-compensation algorithm. Several lines are plotted to see the behavior for different values of actual shift. The aperture is $1.4 \times D_N$ and other parameters are as in Fig. 2.

As might be expected, the curves of BER versus shift measurement error are symmetric when the actual pixel shift is either exactly zero or exactly a half pixel. This makes intuitive sense, since the dependence of BER on shift (Fig. 6) is also symmetric about these points. The only difference is that the BER for zero pixel shift is so low that a significant measurement error is required before the effects are observable. In contrast, the already-high BER at a half-pixel shift grows symmetrically as the error in measuring this shift increases.

For other values of actual shift, Fig. 8 shows a slight but noticeable asymmetry in BER as a function of shift measurement error. For instance, in the range of shifts between 0 and 0.5, it is better to underestimate the shift than to overestimate it. This makes sense since the correction factors applied at 0.5 pixel shift are quite large, meaning that the algorithm changes pixel values by large factors. In contrast, as the measured pixel shift approaches zero (e.g., an actual shift of 0.25, but an error in measuring of -0.25 pixel), the algorithm tends to leave pixel values unchanged, producing a BER that is no worse than the one that would have been obtained without any postprocessing at all. Thus in general, mistakes in estimating the shift that err toward a half-pixel value ($\ldots -1.5, -0.5, 0.5, 1.5, \ldots$) should then increase BER faster than those that lead toward an integer value ($\ldots -2, -1, 0, 1, 2 \ldots$). In any case, mistakes made while measuring the shift cause the largest problems near actual pixel shifts of $|\delta| = 0.5$, where the BER is already large.

6. Wiener Filtering

In this section, we briefly compare the BER improvements provided by the shift-compensation technique against the Wiener filter,[11,13,29] also known as the linear minimum mean-squared-error equalizer.[14,19] Here we follow the nomenclature used in Ref. 14, with the exception that our Gaussian noise is added in the intensity rather than the amplitude domain. Thus in 1-D notation, the nonlinear data channel model is

$$I_k = |(d_k \otimes h_k)^2 + n_k|, \tag{12}$$

where I_k denotes the intensity measured by the kth detector pixel, n_k the additive Gaussian intensity noise, and $\otimes$ indicates the discrete convolution performed between the SLM pixels d_k and the discretized pixel-spread function h_k. The Wiener filter represents the best-case linear filter w_k given this noise level, as quantified by the minimum mean-squared error between the true pixel values d_k and the estimated pixel values,

$$\hat{d}_k = \sum_{j=-N+1}^{N-1} w_j I_{k-j}, \tag{13}$$

summed over $2N + 1$ neighboring pixel values.

As in Reference 14, we obtain the w_k values from

$$w = \mathrm{FT}^{-1}\left[\frac{\mathrm{FT}(R_{dI})}{\mathrm{FT}(R_{II})}\right], \tag{14}$$

where R_{dI} is the crosscorrelation between the input sequence d_k and the output sequence I_k, R_{II} is the autocorrelation of I_k, and FT and FT^{-1} represent the Fourier transform and inverse Fourier transform, respectively. While the above equations describe these operations in one-dimension for clarity, in practice w is a 2-D matrix, from which only the central 5×5 subsection is used as the final convolution kernel.

To obtain such a Wiener filter from our numerical model, we ran the simulated channel on five different random pages (45×45 surrounded by a 3-pixel wide ring of OFF pixels), added random noise, and accumulated the autocorrelation and crosscorrelation matrices R_{II} and R_{dI}. Each of these five data pages was then iterated five times to get a large number of random noise samples as well. Equation (14) was then used to obtain a 5×5 Wiener filter. A new filter was calculated for every different noise value σ_d, aperture D, and pair of misregistrations, δ_x and δ_y. In this way, we generated a bank of Wiener filters that could rapidly respond to changing values of (δ_x, δ_y) using the measured global shift, just as the shift-compensation algorithm does.

Figure 9 shows the resulting improvement in BER offered by the Wiener filter. The parameters for this figure are identical to those used for earlier BER plots: $\sigma_d = 0.05$, contrast of 100:1, and 90% SLM and CCD fill factors. By comparing these curves with those in Fig. 6, it is possible to directly compare shift compensation and Wiener filtering. In fact, many of the curves are almost identical, particularly for shifts along the diagonal $\delta_x = \delta_y$. One point at which Wiener filtering is advantageous is at small apertures and a large shift. In fact, under Wiener filtering the BER at a half-pixel linear shift actually gets worse as the aperture opens up. The reason for this is that when the aperture is small, the broad point-spread function spreads pixel energy over a large number of neighboring pixels. The 5×5 Wiener filter can then obtain a relatively good estimate for the pixel value by drawing upon all these neighbors, despite the nonlinearity of the actual channel. Here also the involvement of a large number of pixels tends to average out the effects of random noise, by helping to further reduce BER. However, the shift-compensation algorithm out-performs Wiener filtering at a half-pixel shift when the aperture is larger. This is a critical point, because this is one of the regimes that dictate the density cost of attaining complete compensation of all possible shifts.

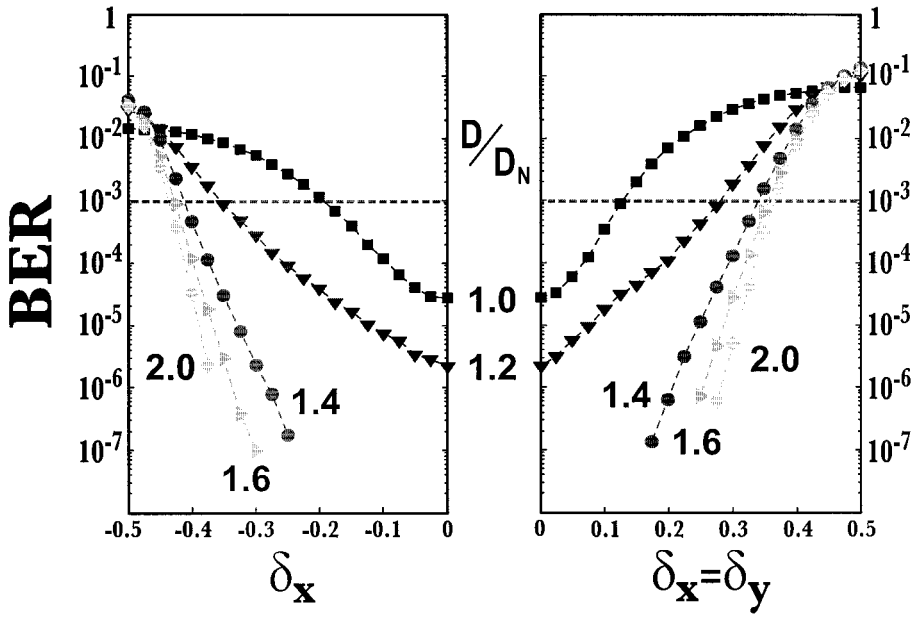

Fig. 9. Simulated BER after Wiener filtering as a function of page shift: (a) along the x axis ($\delta_y = 0$), (b) along the line $y = x$ ($\delta_y = \delta_x$), for various bandlimiting apertures. System parameters are identical to Figs. 2 and 6.

Figure 10 shows this region in closeup, plotting the relevant curves from Figs. 6 and 9. To show the comparison, the horizontal axes on both halves of the plot have been truncated to show just the large shifts ($-0.5 < \delta_x < -0.25$ and $0.25 < \delta_x = \delta_y < 0.5$). We also ran simulations for the true 2-D shift compensation described by Eq. (10), but these curves (not shown) differed little from the 1-D shift compensation case, and at some points were even slightly worse. The advantage of the Wiener filter over shift compensation is that, being a linear filter, the implementation should be more straightforward. The disadvantages are that it offers less performance when the shift is large along one dimension, and that the Wiener filter is already as optimized as a linear filter can be. It also requires many more stored terms than shift compensation: here, given our symmetry assumptions, a 5×5 matrix for each of 20×20 shifts, instead of 3 terms for each of 50 linear shifts (40 for the range $0 < \delta < 1.0$, plus 10 more to handle the averaging near zero, at $-0.25 < \delta < 0$).

A further disadvantage might seem to be that the

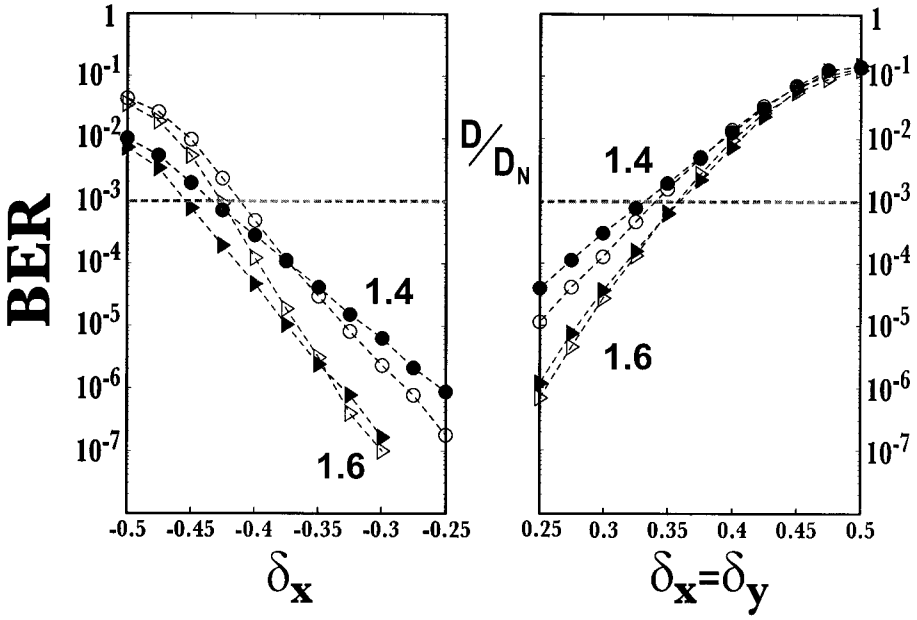

Fig. 10. Comparison of simulated BER for the Wiener filter (open circles and triangles) and for the shift-compensation algorithm (filled circles and triangles). Data are taken from Figs. 6 and 9. Note the truncation of both horizontal axes.

Wiener filter is tightly tied to the particular value of random noise for which it was designed. However, we measured the increase in BER with random noise under two different scenarios: when just a single Wiener filter that was built for the average random noise was used, and when a new Wiener filter was built expressly for each new value of random noise. The difference between the two was minor. So theoretically an enormous bank of Wiener filters should perform almost as well as shift compensation. However, this would probably need to be proven on real experimental data pages, in the same way that shift compensation was demonstrated in Refs. 9 and 27.

7. Discussion

Because it cannot be performed as easily as convolution, shift compensation will always be more complex to implement than the Wiener filter. However, unlike the Wiener filter, there remain several possible improvements to the shift-compensation algorithm. The first would be to improve the performance at a half-pixel shift by simultaneously considering the contributions of other neighboring pixels (those that were assumed to be negligible here). After all, if the present shift-compensation algorithm had precisely inverted the nonlinear channel model, then the histograms after shift compensation for the case of no random noise shown in Fig. 5 would have resembled two delta functions. This illustrates that there remains significant room for improvement in the accuracy of our channel model, especially at a half-pixel shift.

The second and related avenue for improvement is to incorporate the nonlinear channel model described by Eqs. 6 or 10 into an iterative signal-processing scheme.[11–13,17,18] Here estimates for pixels would be updated by recalculating the estimated impact on each pixel from its neighborhood, which in turn would lead to new estimates for all pixels. The discrete channel model described here could be used in such an iterative scheme, with the main complication being the number of coefficients required to handle a large neighborhood and a large number of possible shifts. For instance, rewriting Eq. (10) for a neighborhood of N pixels (counting the target pixel) produces an equation with $N + \binom{2}{N}$ terms. For a 3×3 neighborhood, this is 45 terms; for 5×5, 325 terms. Some of these terms [such as the effect on the central pixel when pixel $(-2, -2)$ and $(+2, +2)$ are both ON] could easily be dropped. However, to cover every possible shift at a granularity of 0.025 pixels, every term deemed important must be represented by a matrix of size 20×20. If the point-spread function cannot be assumed to be symmetric, then this jumps to 40×40. And of course, if the system is space variant, then multiple versions of this large set of numbers would be required.

The third avenue for improving shift compensation is not to fix the poor performance at a half-pixel shift, but to simply average its effects with the good performance at other shifts. We recently implemented this experimentally by simply introducing a small magnification error in the optics.[27] This forces all local shifts to be equally represented at any global page shift, including both regions with good BER performance ($|\delta_x|$, $|\delta_y|$ small) and with poor BER performance ($|\delta_x|$, $|\delta_y|$ approaching 0.5). In addition, the number of algorithm iterations over which error propagation builds up (before reaching a region that is locally registered) is reduced. Using a modified shift compensation algorithm to correct a magnification of 16:15, the detection of near-megapel data pages at constant raw BER $<10^{-3}$ was experimentally demonstrated to be independent of the lateral misalignment.[27] The numerical simulations described in this paper can be used to investigate the susceptibility to random noise, the impact of the choice of magnification error, and the resulting areal density under intentional magnification error, but space constraints require that this be the subject of a subsequent paper.

8. Conclusions

Using a numerical simulation we have investigated the effects of data page misregistration,[9] and the correction of misregistration through shift compensation, on holographic data storage systems. The dependence of BER on the size of the Fourier-transform aperture, the magnitude of the page misregistration in one and two dimensions, and the amount of random noise was investigated. By using the $1/M^2$ dependence of hologram diffraction efficiency on the number of holograms, M, we related the maximum noise tolerance at a given target BER to an effective areal density. As expected, the BER rises and the achievable density falls off as interpixel crosstalk increases, either through a smaller Fourier-transform aperture or larger page misregistration. By extension, the addition of optical distortion or material shrinkage further reduces this effective range, by introducing local shifts within the page that are larger than the overall global misregistration and are thus more prone to error.

We showed that the shift compensation algorithm introduced in Ref. 9 can increase the range of misregistrations that can be tolerated for any given Fourier-transform aperture. The penalty for this is a reduction in the achievable density, because increasing the amount of interpixel crosstalk that is removed by postprocessing inherently increases the susceptibility to random noise. For larger apertures, nearly complete tolerance of all possible shift values can be achieved. The Wiener filter was shown to have similar performance as a function of shift except at the critical points of a half-pixel shift. Future work will investigate the BER and density performance when a small magnification error is intentionally introduced in the optics to provide complete tolerance to misregistration.

Shift compensation can allow system designers to trade off interpixel crosstalk due to global and local pixel misregistration in return for some increased susceptibility to random noise. As with any tradeoff, the more crosstalk removed, the larger the

increased impact of noise. However, up to this point, the only alternative to achieving perfect pixel registration has been gross oversampling at the detector array, leading to density reductions by factors of 4 or more. Thus any effect which causes even a slight translation or distortion of the pixelated data pages has, by necessity, been harshly constrained or avoided. As a result, shift compensation has the potential to open up numerous tradeoffs in holographic data storage, including material shrinkage, thermal expansion, readout at different wavelengths, and readout from moving storage media.

References

1. H. J. Coufal, D. Psaltis, and G. Sincerbox, eds. *Holographic Data Storage* (Springer-Verlag, New York, 2000).
2. 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. Opt. Lett. **26,** 444–446 (2001).
3. S. S. Orlov, Volume holographic data storage. Communications of the ACM **43,** 46–54 (2000).
4. G. W. Burr, E. Mecher, T. Juchem, H. Coufal, C. M. Jefferson, F. Gallego, K. Meerholz, N. Hampp, J. A. Hoffnagle, M. Jurich, R. M. Macfarlane, and R. M. Shelby, "Progress in read-write, fast-access volume holographic data storage," in *Photorefractive Fiber and Crystal Devices: Materials, Optical Properties, and Applications VII, and Optical Data Storage*, S. Yin, F. T. Ya, and H. J. Coufal, eds., Proc. SPIE **4459,** 290–305 (2002).
5. 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," Appl. Opt. **37,** 5377–5385 (1998).
6. M. A. Neifeld and B. L. Zhou, "Optimal pixel profiles for spatially discrete coherent imaging systems," Opt. Commun. **193,** 87–95 (2001).
7. C. Gu, J. Hong, and P. Yeh, "Diffraction properties of momentum-mismatched gratings in photorefractive media," J. Opt. Soc. Am. B **9,** 1473–1479 (1992).
8. 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," Opt. Lett. **25,** 713–715 (2000).
9. G. W. Burr and T. Weiss, "Compensation of pixel misregistration in volume holographic data storage," Opt. Lett. **26,** 542–544 (2001).
10. J. F. Heanue, K. Gurkan, and L. Hesselink, "Signal detection for page-access optical memories with intersymbol interference," Appl. Opt. **35,** 2431–2438 (1996).
11. A. Neifeld, K. M. Chugg, and B. M. King, "Parallel data detection in page-oriented optical memory," Opt. Lett. **21,** 1481–1483 (1996).
12. B. M. King and M. A. Neifeld, "Parallel detection algorithm for page-oriented optical memories," Appl. Opt. **37,** 6275–6298 (1998).
13. K. M. Chugg, X. P. Chen, and M. A. Neifeld, "Two-dimensional equalization in coherent and incoherent page-oriented optical memory," J. Opt. Soc. Am. A **16,** 549–562 (1999).
14. M. Keskinoz and B. V. K. Vijaya Kumar, "Application of linear minimum mean-squared-error equalization for volume holographic data storage," Appl. Opt. **38,** 4387–4390 (1999).
15. V. Vadde and B. V. K. Vijaya Kumar, "Channel modeling and estimation for intrapage equalization in pixel-matched volume holographic data storage," Appl. Opt. **38,** 4374–4386 (1999).
16. A. S. Choi and W. S. Baek, "Equalization for digital holographic data storage," Jpn. J. Appl. Phys. Part 1, **40,** 1737–1740 (2001).
17. W. C. Chou and M. A. Neifeld, "Soft-decision array decoding for volume holographic memory systems," J. Opt. Soc. Am. A **18,** 185–194 (2001).
18. M. Keskinoz and B. V. K. Vijay Kumar, "Efficient modeling and iterative magnitude-squared decision feedback equalization (dfe) for volume holographic storage channels," in *International Conference on Communications* (ICC2001), **9,** 2696–2700 (2001).
19. M. Keskinoz and B. V. K. Vijay Kumar, "Efficient modeling of volume holographic storage channels (VHSC)," in *Optical Data Storage 2000* **4090,** Proc. SPIE, 205–210 (2000).
20. G. W. Burr, W.-C. Chou, M. A. Neifeld, H. Coufal, J. A. Hoffnagle, and C. M. Jefferson, "Experimental evaluation of user capacity in holographic data storage systems," Appl. Opt. **37,** 5431–5443 (1998).
21. 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," Opt. Lett. **22,** 639–641 (1997).
22. J. F. Liu, V. Boopathi, T. C. Chong, Y. H. Wu, and Z. M. Li, "Space-invariant patches in diffraction-limited imaging," Opt. Eng. **39,** 396–400 (2000).
23. J. Hong, I. McMichael, and J. Ma, "Influence of phase masks on cross-talk in holographic memory," Opt. Lett. **21,** 1694–1696 (1996).
24. M.-P. Bernal, G. W. Burr, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, E. Oesterschulze, R. M. Shelby, G. T. Sincerbox, and M. Quintanilla, "Effects of multilevel phase masks on interpixel crosstalk in holographic data storage," Appl. Opt. **36,** 3107–3115 (1997).
25. F. H. Mok, G. W. Burr, and D. Psaltis, "System metric for holographic memory systems," Opt. Lett. **21,** 896–898 (1996).
26. S. Campbell, S. H. Lin, X. M. Yi, and P. C. Yeh, "Absorption effects in photorefractive volume-holographic memory systems. 1: beam depletion," J. Opt. Soc. Am. B **13,** 2209–2217 (1996).
27. G. W. Burr, "Holographic data storage with arbitrarily misaligned data pages," Opt. Lett. **27,** 542–544 (2002).
28. G. W. Burr, T. Weiss, and R. M. Shelby, "Post-processing to correct for optical distortion and material shrinkage in holographic data storage," SPIE Holography newsletter **11,** 4,8 (2000).
29. A. Papoulis, *Probability, random variables, and stochastic processes*, 2nd ed. (McGraw-Hill, New York, 1984).