Optical data storage readout with quadrant pupil detection

T. D. Milster, Z. Chen, E. P. Walker, M. T. Tuell, and E. C. Gage

A novel detection scheme that uses combinations of quadrant signals derived in a pupil of the optical system is described for optical storage devices. The signals arise because of an asymmetry in the reflected light distribution when the focused spot scans data with a nonzero tracking offset. Theoretical and scalar diffraction characterization indicates that the signals may be useful for improved data density by reducing intertrack interference (cross talk). The signals may also be useful for providing a tracking error signal.

Key words: Optical data storage, magneto-optics, high-density recording, cross talk, tracking error signal. © 1996 Optical Society of America

1. Introduction

The most popular form of optical data storage is based on a focused laser beam that scans data on a spinning disk, as shown in Fig. 1. The light source is a laser diode that produces a diverging laser beam. The diverging beam is collimated and circularized. It passes through a beam splitter and is focused by an objective lens onto the recording layer of a spinning disk. Data are stored on the disk in a pattern of spiral tracks. The focused spot follows a track as the disk spins. Compact-disk devices use micrometer-sized pits along the tracks to modulate the reflected light energy and produce a data signal at the detectors. Phase-change devices use crystalline marks on an amorphous background. In erasable magneto-optical data-storage devices, the physical marks and spaces that represent data produce variations in the polarization of the reflected beam. Modulation in the reflected beam that is due to the data pattern is detected with a differential arrangement of silicon detectors that maximizes signal contrast and reduces common-mode noise.[1]

In this paper we describe a technique that can be used to increase mark density by reducing track spacing. The technique can be understood by observing that, as track spacing is reduced, the data signal i_s contains an intertrack interference term Δi in addition to the ideal signal i_0, where i_s is given by

$$i_s = i_0 + \Delta i. \tag{1}$$

The term Δi is commonly referred to as cross talk. We generate a signal that is proportional to the derivative of Δi by a technique we call quadrant pupil detection (QPD). The QPD signal i_ϵ can be subtracted from the derivative of the data signal to yield

$$\frac{\partial i_s{}'}{\partial t} = \frac{\partial i_s}{\partial t} - i_\epsilon \approx \frac{\partial i_0}{\partial t}, \tag{2}$$

which can be used for data detection. Alternatively, i_ϵ can be integrated and subtracted from i_s directly, that is,

$$i_s{}' = i_s - \int_0^t i_\epsilon \mathrm{d}t \approx i_0. \tag{3}$$

In the paragraphs below, we describe in detail how QPD can be applied to optical data-storage systems. In Section 2 we describe the origin of the QPD signal. In Section 3 we discuss results from scalar diffrac-

T. D. Milster, E. P. Walker, and M. T. Tuell are with the Optical Sciences Center, University of Arizona, Tucson, Arizona 85721; Z. Chen is with the State Key Laboratory of Applied Optics, Changchun Institute of Optics and Fine Mechanics, Chinese Academy of Science, Changchun, Jilin Province 130022, China; and E. C. Gage is with the Eastman Kodak Company, KGP Complex, Building 800, 460 Buffalo Road, Rochester, New York 14652-3801.

Received 10 August 1995; revised manuscript received 11 January 1996.

0003-6935/96/142471-06$10.00/0

© 1996 Optical Society of America

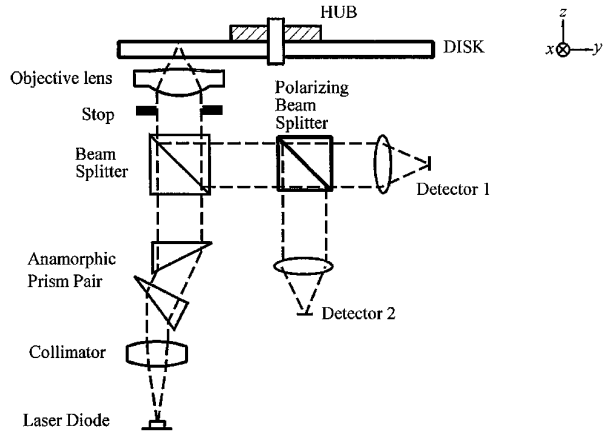

Fig. 1. Layout of a magneto-optical data storage device. A laser diode source is collimated and circularized before passing through a beam splitter. The beam is then focused through a disk substrate. The reflected light is directed to a differential detector arrangement. The hub is used to attach the disk to a spindle motor.

tion modeling and describe two applications of the QPD signal. In Section 4 we analyze noise in the QPD signal. In Section 5 we present results from some preliminary experiments, and in Section 6 we draw conclusions from this work.

2. Theory

The origin of the QPD signal is a diffraction phenomenon present when an off-track spot crosses edges of data marks, as shown in Fig. 2. As the laser spot crosses an edge, an asymmetry is produced in the pupil of the optical system after reflection from the recording layer. (A convenient pupil plane is produced when the stop of the optical system is imaged onto quadrant-cell detectors.) The signal pattern has diagonal symmetry, and it reverses quadrants between leading and trailing edges of the marks. If we divide the pupil of a phase-change or compact-disk system into quadrants labeled a, b, c, and d, the

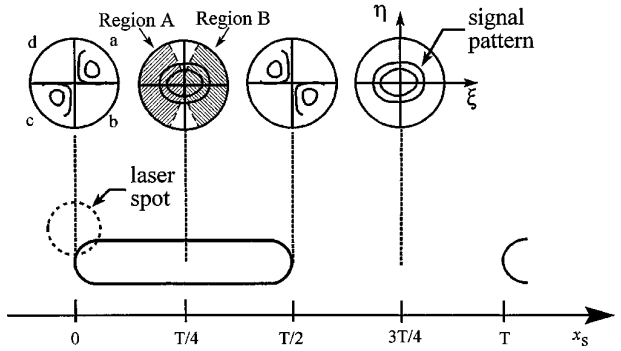

Fig. 2. As the off-track laser spot crosses the edge of a data mark, an asymmetric signal distribution is produced in the pupil of the objective lens. Regions A and B represent overlap of the negative first and positive first diffracted orders, respectively, with the zero order.

data signal is given by

$$i_s = i_a + i_c + i_b + i_d, \tag{4}$$

and the QPD signal is defined by

$$i_\epsilon = C[(i_a + i_c) - (i_b + i_d)], \tag{5}$$

where C is a scaling constant and i_ϵ is a measure of the asymmetry of the signal energy in the pupil.

In a magneto-optic data-storage system, i_ϵ can be realized by reimaging the stop onto quadrant detectors 1 and 2 in Fig. 1. If i_{a1}, i_{a2}, etc., are electrical currents from each quadrant, i_s is given by

$$i_s = (i_{a1} - i_{a2}) + (i_{b1} - i_{b2}) + (i_{c1} - i_{c2}) + (i_{d1} - i_{d2}), \tag{6}$$

and i_ϵ is given by

$$i_\epsilon = C[(i_{a1} - i_{a2}) + (i_{c1} - i_{c2})] - C[(i_{b1} - i_{b2}) + (i_{d1} - i_{d2})]. \tag{7}$$

Note that Eqs. (4) and (5) still apply if we consider $i_a = i_{a1} - i_{a2}$, $i_b = i_{b1} - i_{b2}$, etc.

A similar technique has been proposed for tracking error-signal generation with phase-change media for which only one quadrant detector is necessary.[2] In this geometry, tracks are widely spaced at 1.6 μm. The QPD signal i_ϵ is proportional to y_0, which is the distance between the centroid of the scanning spot and the center of the marks along a track. i_ϵ also reverses sign with y_0, that is, i_ϵ can indicate both the magnitude and direction of the tracking error. By using i_ϵ for tracking error generation, we measure the offset of the scanning spot from the data rather than from the grooves.

We now discuss differences between using QPD with phase-change media and with magneto-optic media. With phase-change media, the QPD signal manifests itself as redistribution of the irradiance in the pupil. The resulting QPD signal can be quite large. For magneto-optic media, data marks can be considered as quarter-wave-deep pits that reflect a small-signal component from the recording layer.[3] The signal component is polarized in a direction that is orthogonal to the incident polarization. Because the signal component is small (typically ~1% of the reflected energy), the QPD signal is also small. Therefore it is important to study the signal and noise properties of i_ϵ.

The origin of the QPD signal can be explained with a simplified description of the readout geometry. Figure 3 illustrates the plane of the recording layer, with the focused spot scanning in the x direction along $y = 0$. A data track is offset from the scanning spot by an amount y_0. The pattern of marks and spaces is represented by $m(x, y)$, where

$$m(x - x_s, y) = \frac{1}{2}\cos\left[\frac{2\pi}{T}(x - x_s)\right]\delta(y - y_0) + \frac{1}{2}, \tag{8}$$

and x_s is the scan coordinate that represents the in-track displacement of the pattern. Note that Eq.

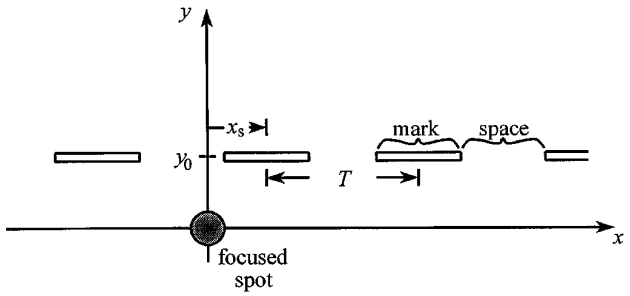

Fig. 3. Hypothetical geometry for deriving QPD characteristics. Thin marks are offset by an amount y_0 from the scanning spot.

(8) is a narrow cosinusoidal amplitude pattern on a constant background. A magneto-optic data pattern requires a more complicated description, but the amplitude pattern in Eq. (8) is sufficient to illustrate the QPD concept. For simplicity, we assume that the stop is a rectangle defined by

$$P(\xi, \eta) = \text{rect}(\xi/a)\text{rect}(\eta/a), \tag{9}$$

where ξ and η are scaled aperture coordinates given by $\xi = x'/\lambda f$ and $\eta = y'/\lambda f$, f is the focal length of the objective lens, and (x', y') represent the real aperture coordinates.

The reflected beam from the recording layer consists of a zero order and positive and negative first diffracted orders from the cosine term in Eq. (8). The angular separation of the positive and negative first diffracted orders increases with data frequency $1/T$. The phase of the zero order is uniform and does not depend on scan position or offset. The positive and negative first diffracted orders contain linear phase terms that are due to offset y_0. They also contain a constant phase term that depends on the scan coordinate x_s. Mathematically, the amplitude and phase distribution in the ξ direction is given by

$$S(\xi) = \tfrac{1}{4}[2\delta(\xi) + \exp(-2\pi i x_s/T)\delta(\xi - 1/T) + \exp(2\pi i x_s/T)\delta(\xi + 1/T)] * \text{rect}(\xi/a), \tag{10}$$

where the asterisk denotes convolution. The amplitude and phase distribution of the positive and negative first orders in the η direction is given by

$$S(\eta) = \exp(-2\pi i y_0 \eta) * \text{rect}(\eta/a). \tag{11}$$

The convolution in Eq. (11) can be determined in closed form because of the symmetry of $P(\xi, \eta)$. The result is

$$S(\eta) = \frac{\sin \pi y_0 a}{\pi y_0} \exp(-2\pi i y_0 \eta). \tag{12}$$

Note that $S(\eta)$ is a constant multiplied by a linear phase factor. Region A is defined as the area where the negative first order and the zero order overlap, as shown in Fig. 2 for a circular stop. The linear phase factors of Eqs. (10) and (12) combine to yield an irradiance pattern in region A given by

$$I_A(x_s, \eta; y_0) \propto 1 + |c_A|^2 + 2|c_A|\cos 2\pi(y_0\eta - x_s/T), \tag{13}$$

where c_A is a constant. Likewise, region B is defined as the area where the positive first order and the zero order overlap, as shown in Fig. 2. In region B the irradiance pattern is given by

$$I_B(x_s, \eta; y_0) \propto 1 + |c_B|^2 + 2|c_B|\cos 2\pi(y_0\eta + x_s/T). \tag{14}$$

Expressions (13) and (14) describe simple cosinusoidal fringes in the plane of the aperture. The spatial frequency of the fringes in the η direction is inversely proportional to the offset y_0. Constructive interference (e.g., a bright fringe) occurs under the condition

$$y_0\eta \pm x_s/T = j, \tag{15}$$

where j is an integer, the minus sign refers to region A, and the plus sign refers to region B. As the spot scans the disk, x_s increases, and Eq. (15) indicates that the bright fringe moves in the direction of increasing η for region A and in the direction of decreasing η for region B. When the spot is centered on a mark or a space, $x_s = mT/2$, where m is an integer, and the resulting fringe pattern is symmetric in regions A and B. When $x_s \neq mT/2$, the pattern is asymmetric. Maximum asymmetry occurs when $x_s = (m/2 + 1/4)T$, that is, when the spot is crossing an edge.

The frequency response of the QPD signal is an interesting phenomenon. Because the QPD signal is derived from edges of marks, it is similar to a push–pull tracking signal. Push–pull signals have a frequency response maximized near NA/λ, where NA is the numerical aperture of the objective lens and λ is the wavelength.[4] However, a true frequency-response characteristic is not easily described for the QPD signal. The QPD signal is based on diffraction from a two-dimensional edge, so a simple cosine decomposition does not adequately describe the problem. In any case, the QPD signal response is different from the i_0 frequency response.

Real data patterns are more complicated than the simple description in Eq. (8). However, the concepts presented in this section can be used to provide a heuristic explanation of the phenomenon. For example, the mark patterns are typically 0.8 μm wide in the y direction, and the corresponding form of Eq. (12) is not a simple constant multiplied by a linear phase factor. In reality, the differences do not dramatically affect the fringe pattern. The fringe shift is observed as predicted from Eq. (15). The maximum asymmetry is present when the focused spot covers the mark edge, and the polarity of the QPD signal changes as y_0 goes through 0. In Section 3 we show that, with realistic mark and space patterns, i_ϵ is proportional to the derivative of the off-track signal Δi.

3. Scalar Diffraction Modeling

The QPD signal is simulated with a scalar diffraction-modeling program that is designed to analyze magneto-optical disk systems.[5] Two-dimensional matrices are used to describe laser beam profiles and mark patterns. The recording-layer matrix contains patterns of 0.72-μm-wide marks arranged in tracks. Track spacings as well as the mark and space widths are adjustable. Mark edges are rounded to simulate realistic mark shapes. At the objective lens, the Gaussian-beam $1/e^2$ diameter fills 90% of the aperture. Other model parameters include $\lambda = 780$ nm and NA = 0.5. The power illuminating the recording layer is 2.0 mW. The Kerr angle is 0.25°, and the medium exhibits no ellipticity.

Propagation of the laser beam through the system starts at the plane of the stop. The complex field matrix that represents the aperture is Fourier transformed to calculate the beam matrix incident upon the recording layer. Multiplication of the incident-beam matrix and the recording-layer matrix yields the beam matrix just after reflection, which is inverse Fourier transformed to calculate the reflected field matrix in the plane of the stop. After a correction for the beam splitter, the signal current i_s and the QPD current i_ϵ are calculated by operation on irradiance values in pixels that correspond to i_a, i_b, i_c, and i_d. It is necessary to keep track of the polarization at each pixel because the information on the magneto-optic recording layer is detected as polarization modulation in the reflected beam.

Signal current i_s is shown in Fig. 4(a) for a mark pattern that consists of three different-sized marks, as shown above the signal trace. The optical spot is scanning at $y_0 = 0.8$ μm above the center of the track. The longest mark and space correspond to a temporal frequency of 1 MHz when the medium velocity is 6.6 ms^{-1}. Likewise, the three following marks and spaces correspond to 3 MHz, and the last six marks and spaces correspond to 6 MHz. The amplitude of i_s decreases with increasing frequency because of the rolloff of the optical transfer function.[3] The derivative $\partial i_s/\partial x_s$ is shown in Fig. 4(b) for the same mark pattern. Figure 4(c) displays the QPD signal i_ϵ with $C = 1$. The shape of i_ϵ is similar to the shape of Fig. 4(b), except that the amplitude of i_ϵ does not have the same frequency dependence. To make the shapes more nearly equal, an electronic filter with response $A(f)$, where f is the electrical frequency, can be applied to i_ϵ to attenuate the high frequencies. When amplifier gains are adjusted properly, the filtered version of i_ϵ and the signal proportional to $\partial i_s/\partial x_s$ are nearly identical, as shown in Fig. 5.

Results of the previous paragraph can be applied to increasing the storage density of an optical data-storage device. Consider the track geometry shown in Fig. 6, where there are no grooves on the recording surface. Track n is being scanned with the focused spot. Tracks $n - 1$ and $n + 1$ are spaced 0.8 and 1.6 μm, respectively, from track n. If this pattern of tracks is repeated over the entire disk, the mark and space density increases by a factor of 1.33 because of the increased track density. As the spot scans with increasing x_s, the intertrack interference Δi from track $n - 1$ is significant, and i_s is distorted, as indicated in Eq. (2). Intertrack interference from track $n + 1$ is insignificant because it is relatively far away from track n, but the effect is included in our analysis. Because the spot is scanning along the center of track n, the only significant QPD signal is due to track $n - 1$. Therefore $i_\epsilon \propto \partial \Delta i/\partial x_s$, and

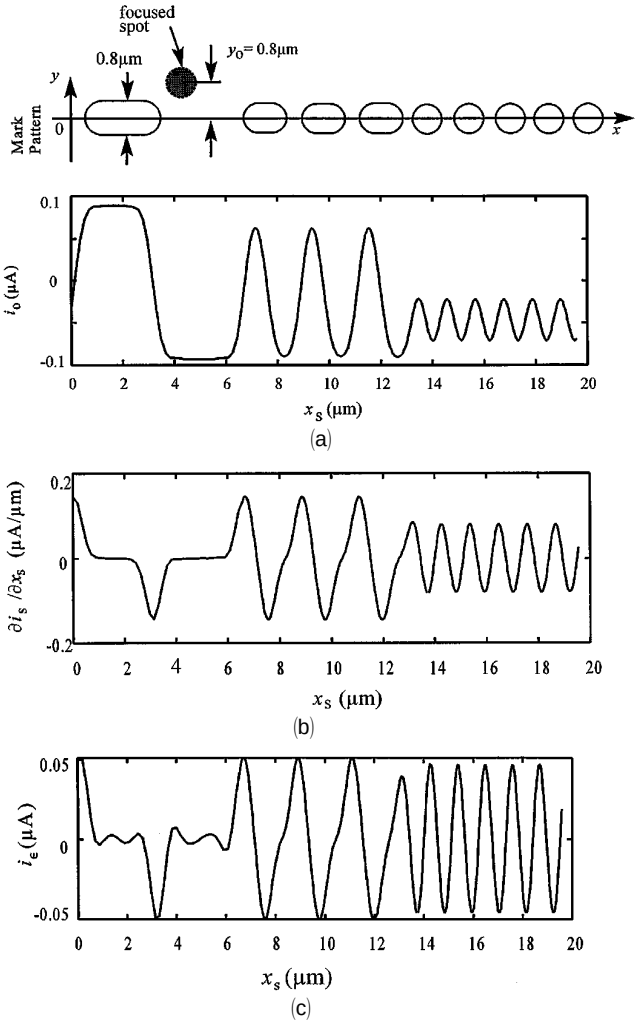

Fig. 4. Readout signals versus scan position: (a) usual data signal is i_s, (b) $\partial i_s/\partial x_s$, and (c) QPD signal i_ϵ.

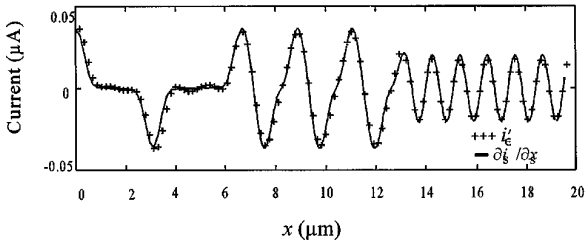

Fig. 5. Comparison of the filtered i_ϵ and that of Fig. 4(c).

intertrack interference can be canceled by the subtraction of $\int i_\epsilon dx_s$ from i_s. Relative signal currents displayed in Fig. 6 show effects of the subtraction. The curve made up of open circles displays the uncorrected signal i_s. Worst distortion is present when the high-frequency pattern of track n is next to a low-frequency mark or space on track $n - 1$. The distortion takes the form of a signal offset. The solid curve is the ideal signal i_0. The curve comprising crosses is the corrected signal $i_s - \int i_\epsilon dx_s$. Modeling results indicate that the difference between the ideal and the corrected signals is less than 3% of the ideal signal under the worst-case conditions, which include ± 0.05 μm of track error for the spot following track n and ± 1.0 μm of defocus. This compares favorably with the 1%–2% residual cross talk from track $n + 1$.

QPD can also be used as a tracking error signal that is proportional to offset y_0 of the spot from track center. Consider a flat disk with no groove pattern. Data tracks are separated by 1.6 μm. Instead of using the QPD signal i_ϵ to cancel intertrack interference, it is used to measure y_0. In Section 2 it was shown that the QPD signal is equal to 0 when the spot is centered on the track, i.e., $y_0 = 0$. i_ϵ increases toward higher positive values for leading edges of marks as y_0 increases. Conversely, i_ϵ decreases toward more negative values for leading edges of marks as the spot moves farther below the track, i.e., as y_0 becomes more negative. Figure 7 illustrates the peak of the QPD signal as a function of the off-track position y_0, which exhibits the bipolar characteristic necessary for a tracking error signal.

4. Noise Analysis

There are several important components of noise. These include shot noise, electronic noise, disk noise, and laser noise. Shot noise arises because of the random arrival and detection of photons at the detectors and is the same magnitude for QPD as for normal differential detection. Electronic noise is a function of the particular components chosen to implement the circuitry. Laser noise and disk noise are usually the most significant noise sources.

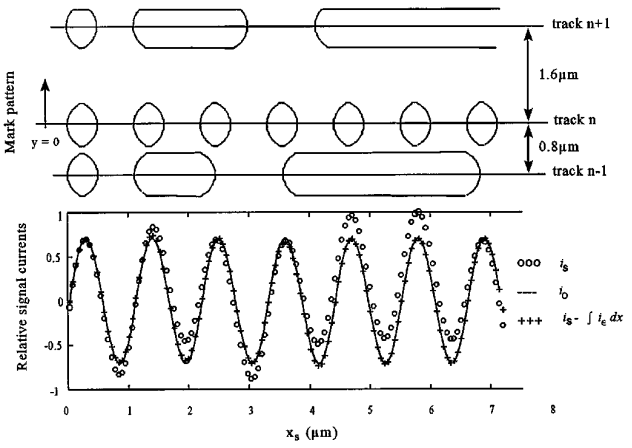

Fig. 6. Implementation with a 33% improvement in data density. The integral of the QPD signal $\int i_\epsilon dx_s$ is used to subtract the interference from the closely spaced track.

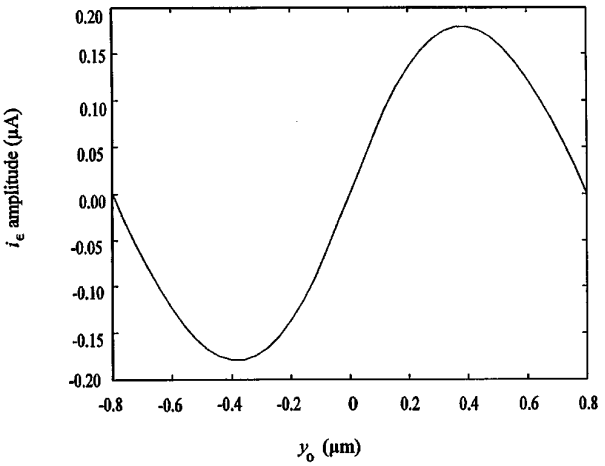

Fig. 7. QPD signal used as a tracking error signal. The peak of i_ϵ is displayed versus the track offset y_0.

The behavior of laser and disk noises in the stop is different from that of shot or electronic noises. Electronic noise varies with frequency, but it is not affected by light level and adds a constant background noise to the signal. Root-mean-square (rms) noise current that is due to shot noise is given by

$$\langle i_{\text{shot}} \rangle \propto (I_a + I_c + I_b + I_d)^{1/2}, \tag{16}$$

where I_a, I_b, I_c, and I_d are the dc electrical currents from each set of quadrants and $\langle \cdots \rangle$ denotes the rms value. Contributions to shot noise from different parts of the detectors add in quadrature because they are not correlated in time. Laser noise and disk noise are distributed in a circularly symmetric pattern in the plane of the stop. Because the detectors are located at an image plane of the stop, the disturbance at any instant in time in one quadrant is correlated with the disturbance in the other quadrants. For example, a laser power fluctuation that is due to mode hopping could momentarily change the light level emitted from the laser. The change in light level affects each quadrant at the same instant in time. Similar arguments can be made with respect to disk noise.

Effects of laser noise and disk noise in the QPD signal can be described by consideration of the differential noise components δi_a, δi_b, δi_c, and δi_d on each quadrant of the detectors. For phase-change and pit-type media, QPD noise current δi_ϵ is given by

$$\delta i_\epsilon = (\delta i_a + \delta i_c) - (\delta i_b + \delta i_d), \tag{17}$$

which is similar to the QPD signal current given by Eq. (6). With Eq. (17), we assume that the noise components are correlated in time. If laser and disk noise exhibit circularly symmetric distribu-

tions, $\delta i_\epsilon = 0$. Our experiments suggest that the noise distributions are not entirely symmetric, so δi_ϵ is small but not 0.

The signal-to-noise ratio (SNR) γ_ϵ in decibels for the QPD signal is given by

$$\gamma_\epsilon = 20 \log_{10}\left(\frac{\langle i_\epsilon \rangle}{\langle \delta i_\epsilon \rangle}\right), \tag{18}$$

where $\langle i_\epsilon \rangle$ is the rms QPD current and $\langle \delta i_\epsilon \rangle$ is the rms QPD noise current in a 6-MHz bandwidth. Likewise, the SNR for the single-track signal current i_0 is given by

$$\gamma_0 = 20 \log_{10}\left(\frac{\langle i_0 \rangle}{\langle i_n \rangle}\right), \tag{19}$$

where $\langle i_n \rangle$ is the total rms noise current after differential detection in a 6-MHz bandwidth. For commercial optical data storage devices, $\gamma_0 \geq 20$ dB, which indicates that $\langle i_0 \rangle \geq 12.6 \langle i_n \rangle$. Experimentally we find that $\langle i_\epsilon \rangle \approx \langle i_0 \rangle / 10$. It is reasonable to assume that, in the presence of an asymmetry in the noise distribution, $\langle \delta i_\epsilon \rangle = \alpha \langle i_n \rangle$, where α is a constant. A combination of these relations gives

$$\gamma_\epsilon \geq 20 \log_{10}\left(\frac{1.26}{\alpha}\right). \tag{20}$$

If $\alpha = 0.1$, then $\gamma_\epsilon = 22$ dB. For magneto-optical devices, the differential-detection technique described in Eq. (7) cancels most common-mode laser and disk noise. $\langle i_n \rangle$ and $\langle \delta i_\epsilon \rangle$ are dominated by noise components that pass through the differential channel. Therefore we do not expect that the noise cancellation described by Eq. (17) improves the SNR dramatically for magneto-optic systems.

5. Experiment

The system shown in Fig. 1 is implemented with an air-bearing spindle to rotate a grooveless disk. An NA = 0.5 objective lens focuses light from a $\lambda = 785$ nm laser diode through a 1.2-mm glass coverplate onto the magneto-optic recording layer. Focus control is maintained through a closed-loop focus servo and a voice-coil actuator that moves the objective lens in the focus direction. Radial runout of the rotary air-bearing spindle is less than 0.02 μm, so a tracking servo is not necessary.

A 4.0-MHz pattern of equal-width marks and spaces was recorded. The medium velocity was 6.6 ms^{-1}. The data signal i_s and the QPD signal i_ϵ were measured as the focused spot was shifted off track. Time-domain signals i_ϵ and i_s are shown in Fig. 8. The off-track distance is $y_0 = 0.5$ μm. Also shown in Fig. 8 is a calculated derivative di_s/dx_s from the digitized data. Figure 8 indicates that i_ϵ is a good approximation to di_s/dx_s.

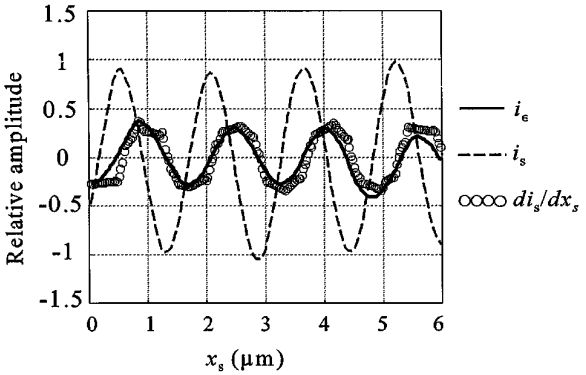

Fig. 8. Data signal i_s, QPD signal i_ϵ, and derivative di_s/dx_s for a pattern of closely spaced data marks. The off-track distance $y_0 = 0.5$ μm.

6. Conclusions

We show theoretically and with scalar diffraction modeling that the distribution of the signal pattern in the pupil of a magneto-optic data-storage device is asymmetric when the focused spot scans a data track with nonzero offset. The maximum asymmetry is present when the spot aligns with a mark edge. The asymmetry is detected with a simple arrangement of data detectors, and the QPD signal is proportional to the derivative of the data signal. The QPD signal does not have the same frequency dependence as the data signal. An electrical filter can be used to attenuate the QPD signal so that it more nearly matches the derivative of the off-track data signal. The filtered QPD signal can then be used to subtract intertrack interference created by closely spaced tracks. An improvement of density by a factor of 1.33 can be realized in a straightforward case. The unfiltered QPD signal can also be used to provide a bipolar tracking error signal. Our experimental results verify the form of the QPD signal.

The authors acknowledge financial support from the Optical Data Storage Center at the University of Arizona, Eastman Kodak, Inc., and the Graduate Assistance in Areas of National Need program.

References

1. D. Treves and D. S. Bloomberg, "Signal, noise and codes in optical memories," Opt. Eng. **25,** 881–891 (1986).
2. T. Shibata, "Signal detection system for use in an optically reproducing apparatus," U.S. patent 4,467,462 (21 August 1984).
3. T. D. Milster and C. Curtis, "Analysis of superresolution in magneto-optic storage devices," Appl. Opt. **31,** 6272–6279 (1992).
4. G. Brouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. Schouhamer Immink, *Principles of Optical Disk Systems* (Hilger, Accord, Mass., 1985), p. 22.
5. J. P. Treptau, T. D. Milster, and D. G. Flagello, "Laser beam modeling in optical storage systems," in *Modeling and Simulation of Laser Systems II,* A. D. Schnurr, ed., Proc. SPIE **1415,** 317–321 (1991).