

Analysis of Nonlinear Transition Shift and Write Precompensation in Perpendicular Recording Systems

Zheng Wu, *Student Member, IEEE*, Paul H. Siegel, *Fellow, IEEE*, Jack K. Wolf, *Fellow, IEEE*, and H. Neal Bertram, *Fellow, IEEE*

Abstract—In high density perpendicular magnetic recording channels, nonlinear transition shift (NLTS) is one of the distortions that can degrade the system performance. Write precompensation is a standard method used to combat the negative effect of NLTS. In this paper, we present an analysis of the bit-error-rate (BER) for perpendicular recording systems with NLTS and write precompensation. Media jitter noise and additive white Gaussian noise are also considered in the model. A BER lower bound is derived, as well as a more easily computed estimate of the bound. The write precompensation values that numerically minimize the estimate of the BER lower bound prove to be very close to those found using Monte-Carlo channel simulation. We then apply these methods to the design of multilevel precompensation schemes, for which the optimization of precompensation values by Monte-Carlo channel simulation is computationally infeasible. The results show that for higher recording densities subject to increased ISI and noise, the use of more complex precompensation schemes does not significantly improve the system performance.

Index Terms—NLTS, write precompensation, jitter noise, perpendicular recording

I. INTRODUCTION

In a high density perpendicular recording system, nonlinear effects can distort the read-back signal and degrade the system performance. Nonlinear transition shift (NLTS) induced by demagnetization from previously written transitions is one example. As in longitudinal recording, the NLTS in a perpendicular recording channel can be measured by time or frequency analysis of the read-back signal corresponding to a carefully chosen input data pattern [1], [2]. The distortion caused by NLTS can be reduced by the use of write precompensation, whereby, for specific data patterns, deterministic offsets are added to the timing of written transitions. A simple and commonly used precompensation scheme is dibit precompensation, which affects the second transition of a pair of adjacent transitions. In practice, the timing offsets in write precompensation are optimized empirically in order to minimize the bit-error-rate (BER).

Z. Wu was with the Center for Magnetic Recording Research, University of California, San Diego, CA 92093, USA. She is now with Link-A-Media Devices, Santa Clara, CA 95051.

P. H. Siegel, J. K. Wolf and H. N. Bertram are with the Center for Magnetic Recording Research, University of California, San Diego, CA 92093, USA.

H. N. Bertram was also with Hitachi Global Storage Technologies, San Jose, CA 95135, USA. He is now with Western Digital Corporation, San Jose, CA 95138

Email: {z2wu,psiegel,jwolf,nbertram}@ucsd.edu

There are very few theoretical results on optimal precompensation of NLTS in recording channels to minimize the BER because of the complex nature of the nonlinear effects. Lim and Kavčić [3] presented a dynamic programming method to optimize write precompensation for a longitudinal recording channel with partial erasure, NLTS and additive white Gaussian noise (AWGN). Their objective was to minimize the mean-squared error (MSE) between the output signal of the noisy, nonlinear channel model and that of the noiseless, linear channel model, rather than to minimize BER. They allowed the use of a different precompensation value for each transition. The optimization procedure and the resulting precompensation scheme in [3] would be too complex to implement in a real system. In practice, it is typical to use a small number of different precompensation values corresponding to a specified subset of data patterns. (We refer to a scheme with more than one such precompensation value as a multilevel scheme.) In a previous work [4], we compared the precompensation values obtained by minimizing the BER to those obtained by minimizing the MSE for two specific precompensation schemes. The values were close, though not identical.

In this paper, we will present an analysis of the BER for systems with NLTS and write precompensation. A lower bound on the BER is derived, as well as a more easily computed estimate of the lower bound. We evaluate these numerically, and compare their BER performance predictions to the results of Monte-Carlo simulation. We find that the optimal precompensation values obtained using these BER estimates are very close to those found using the Monte-Carlo method. This motivates the application of similar analytic techniques to the optimization of more complex multilevel precompensation schemes, for which Monte-Carlo simulation is computationally impractical.

The paper is organized as follows. Section II presents the channel model and defines the order-1 and order-2 channel approximations that we use in our analysis and simulations. Section III gives the derivation of the pairwise error event probability for the two channel approximations, as well as upper and lower bounds on the BER. An easily computed estimate of the lower bound is also presented. Section IV shows BER results obtained by numerical evaluation of the lower bound estimate for two simple precompensation schemes. These performance results, as well as the optimized precompensation values, are compared to results found by Monte-Carlo simulation. In Section V, we use the BER analysis to

gain insight into the observed performance benefits of write precompensation. We then extend the application of the BER estimation techniques to the optimization of more elaborate multilevel precompensation algorithms. Section VI concludes the paper.

II. CHANNEL MODEL AND WRITE PRECOMPENSATION

We consider a channel model with NLTS, jitter noise and AWGN, the same as in [4]. Let the channel transition response be

$$s(t) = V_{\max} \operatorname{erf}\left(\frac{0.954t}{T_{50}}\right), \quad (1)$$

where $\operatorname{erf}(\cdot)$ is the error function defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

and T_{50} is the width of the transition response from $-V_{\max}/2$ to $V_{\max}/2$.

Let $\{x_i\}$ be the input binary data sequence to the channel, $x_i \in \{-1, +1\}$. The transition sequence induced is thus $d_i = \frac{x_i - x_{i-1}}{2}$, $d_i \in \{-1, 0, +1\}$. The channel output $z(t)$ can be written as

$$z(t) = \sum_i d_i s(t + \delta_i + a_i - iB) + n_W(t). \quad (2)$$

Here, δ_i is the net shift of the transition d_i with respect to its nominal location in the recording medium, a_i is the random position jitter for transition d_i , B is the channel bit spacing (as well as the sampling period), and $n_W(t)$ is the electronics noise. For $d_i = 0$, we set $a_i = 0$, whereas for $d_i \neq 0$, a_i is a zero mean Gaussian random variable with variance σ_j^2 . The jitter values for recorded transitions are mutually independent. The electronics noise $n_W(t)$ is modeled as a zero-mean, AWGN process. The variance of the discrete time AWGN samples $n_W(kB)$ is denoted by σ_W^2 . We define the signal-to-AWGN ratio to be $SNR_W = 10 \log_{10}(V_{\max}^2/\sigma_W^2)$. This SNR definition is taken from [5] where it was introduced in order to facilitate the study of the separate effects of jitter noise and AWGN in channels with NLTS and precompensation.

We can write the net shift as $\delta_i = \tau_i + \Delta_i$, where τ_i is the NLTS induced by previously recorded transitions, and Δ_i is the precompensation value for the transition d_i . By convention, for $d_i = 0$, we set $\delta_i = \tau_i = \Delta_i = 0$. According to the model proposed by Bertram and Nakamoto [6], [7], the NLTS of a transition is determined by the distance from the writing location of the current transition to the actual locations of previously written transitions, assuming a fixed head-media configuration. Therefore, τ_i is a function of the transition sequence d_0, \dots, d_i , the net shifts of the previously recorded transitions $\delta_0, \dots, \delta_{i-1}$, and Δ_i . In perpendicular recording, the NLTS always shifts a transition away from previous transitions. Referring to (2), we see that τ_i is non-positive and the precompensation value Δ_i is non-negative. Given a head-media configuration, a transition sequence and a precompensation scheme, the net transition shifts of all the recorded transitions are uniquely determined.

The complexity of channel simulation can be reduced if we approximate the channel output by truncating the Taylor

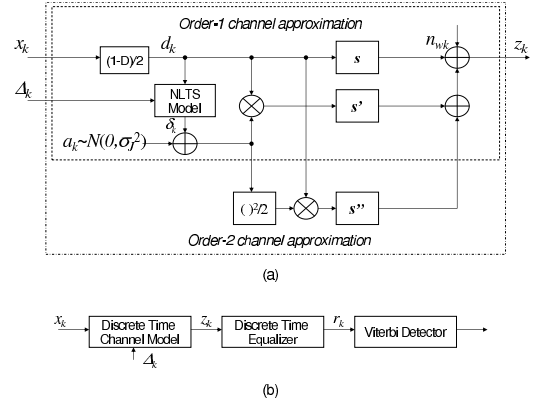


Fig. 1. (a) Discrete time channel model; (b) Discrete time system diagram

expansion of the transition response. The *order-1* channel approximation is obtained by considering terms up to the first derivative term in the expansion of the transition response:

$$z(t) \approx \sum_i x_i h(t - iB) + \sum_i d_i (\delta_i + a_i) s'(t - iB) + n_W(t), \quad (3)$$

where $h(t) = (s(t) - s(t - B))/2$ is the dipulse response. The *order-2* channel approximation takes into account both the first and second derivatives terms:

$$z(t) \approx \sum_i x_i h(t - iB) + \sum_i d_i (\delta_i + a_i) s'(t - iB) + \sum_i d_i \frac{(\delta_i + a_i)^2}{2} s''(t - iB) + n_W(t). \quad (4)$$

A discrete time channel model and system diagram are shown in Fig. 1. The discrete time channel output is passed through the equalizer before it enters the sequence detector. In this paper, we consider a Viterbi detector that is matched to the equalizer output target.

In practice, recording systems usually use only a small number of precompensation levels corresponding to selected patterns of recently recorded transitions. These transitions play the most significant role in determining the NLTS value of the transition being written. In [4], we considered two precompensation schemes: the so-called dibit precompensation scheme, as well as a two-level scheme. The dibit scheme considers only the most recently recorded bit, and applies precompensation only to the second transition in a pair of adjacent transitions, i.e., $\Delta_i = \Delta$ only if $d_i \neq 0$ and $d_{i-1} \neq 0$, otherwise, $\Delta_i = 0$. The dibit precompensation scheme can be thought of as a one-level precompensation. The two-level precompensation scheme considers two configurations of transitions in the two preceding bit positions, namely a transition only in the most recent position or transitions in both of the preceding positions. In this paper, we examine two more precompensation schemes that take into account all possible bit configurations in the preceding two bits or preceding three bits. We refer to these as the 2-bit look-back scheme and 3-bit look-back scheme, respectively. The 2-bit

look-back scheme applies possibly different precompensation values to three distinct bit patterns, while the 3-bit look-back scheme may use up to seven different values for the seven distinct bit patterns. Of course, as we increase the number of patterns to which precompensation is applied, it becomes more difficult to find the optimal set of precompensation values by numerical optimization using Monte-Carlo simulation.

III. BIT-ERROR-RATE ANALYSIS

The performance analysis of a Viterbi detector for an ISI channel can be found in several references, such as [8], [9]. The analysis has been extended to magnetic recording channels with jitter noise and modified versions of the Viterbi algorithm [10]–[12]. In this section, we develop an estimate of the BER at the Viterbi detector output when NLTS is present. We first derive the pairwise error probability corresponding to a given error event. Then, using various union bounds, we develop upper and lower bounds on the BER. A more easily computed estimate of the lower bound is then introduced.

A. Pairwise Error Probability Analysis

Let $\mathbf{x}^{(1)}$ denote the recorded sequence and $\mathbf{x}^{(2)}$ the detected sequence. We denote by ξ_1 and ξ_2 the paths in the Viterbi detector trellis that correspond to $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, respectively. We say that ξ_1 is the correct path and ξ_2 is the incorrect path. We assume that ξ_1 and ξ_2 diverge at time k and remerge at time $k + M$. We refer to M as the length of the error event.

The *pairwise error probability* of the error event $\epsilon_M = (\xi_1 \rightarrow \xi_2)$ is the probability that ξ_2 is chosen as the detected path instead of ξ_1 , given that $\mathbf{x}^{(1)}$ is the written sequence. We should note that in our case, the pairwise error probability for event $(\xi_1 \rightarrow \xi_2)$ may differ from that for event $(\xi_2 \rightarrow \xi_1)$ because of the data-dependent noise and NLTS.

An error event occurs when the accumulated metric on the incorrect path is smaller than that on the correct path. In a conventional Viterbi detector, the Euclidean distance is used as the branch metric. Therefore, the pairwise error probability can be expressed as follows:

$$\Pr(\xi_1 \rightarrow \xi_2 | \mathbf{x}^{(1)}) = \Pr\left(\sum_{i=k}^{k+M-1} (r_i - y_i^{(1)})^2 > \sum_{i=k}^{k+M-1} (r_i - y_i^{(2)})^2 | \mathbf{x}^{(1)}\right), \quad (5)$$

where r_i is the equalized channel output sample at time i when $\mathbf{x}^{(1)}$ is transmitted, and $y_i^{(1)}$ and $y_i^{(2)}$ are the noiseless channel outputs at time i corresponding to the input data sequences $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, respectively. Assume that the equalization target of the system is $g(D) = g_0 + g_1 D + \dots + g_J D^J$, where D is a unit delay operator. We can then write the branch output labels as $y_i^{(1)} = \sum_{l=0}^J g_l x_{i-l}^{(1)}$ and $y_i^{(2)} = \sum_{l=0}^J g_l x_{i-l}^{(2)}$.

The noise value at time i is given by $n_i = r_i - y_i^{(1)}$. After some straightforward calculation, we can express the pairwise error probability as

$$\Pr(\xi_1 \rightarrow \xi_2 | \mathbf{x}^{(1)}) = \Pr\left(\sum_{i=k}^{k+M-1} 2n_i(y_i^{(1)} - y_i^{(2)}) < -\sum_{i=k}^{k+M-1} (y_i^{(1)} - y_i^{(2)})^2 | \mathbf{x}^{(1)}\right). \quad (6)$$

In the above equation, the outputs $y_i^{(1)}$ and $y_i^{(2)}$ are completely determined by the data sequence $\mathbf{x}^{(1)}$ and the specific error event. Therefore, the pairwise error probability is equal to the probability that a linear combination of the noise samples $n_i, i = k, \dots, k + M - 1$ is smaller than a deterministic value. If we can find the distribution of the linear combination of the noise samples, the probability can be calculated directly. We now discuss how to calculate this probability for order-1 and order-2 approximations of the channel with NLTS.

1) *Order-1 Channel Approximation*: In the order-1 channel approximation, the noise at time i can be written as

$$n_i = \sum_j x_{i-j}^{(1)} \tilde{h}_j - \sum_j x_{i-j}^{(2)} g_j + \sum_j d_{i-j}^{(1)} (\delta_{i-j} + a_{i-j}) \tilde{s}'_j + w_i, \quad (7)$$

where \tilde{h}_j and \tilde{s}'_j denote the convolution of the FIR equalizer taps with the samples of the channel dipulse response and the first derivative of the transition response, respectively. The equalized sample of the AWGN at time k is denoted by w_k .

In perpendicular recording systems, \tilde{h}_j and \tilde{s}'_j will vanish when j goes to $+\infty$ and $-\infty$. Therefore, we can assume that sequences $\{\tilde{h}_j\}$ and $\{\tilde{s}'_j\}$ have finite length. Thus, $n_i, i = k, \dots, k + M - 1$ are non-zero mean, jointly distributed Gaussian random variables given $\mathbf{x}^{(1)}$. Equation (6) can be calculated by an evaluation of the Q -function, the tail probability of the standard Gaussian density:

$$\Pr(\xi_1 \rightarrow \xi_2 | \mathbf{x}^{(1)}) = Q\left(\frac{\boldsymbol{\lambda}^T \boldsymbol{\lambda} + 2\boldsymbol{\lambda}^T \boldsymbol{\mu}}{2\sqrt{\boldsymbol{\lambda}^T \boldsymbol{\Sigma} \boldsymbol{\lambda}}}\right) \quad (8)$$

where column vector $\boldsymbol{\lambda}$ is $(y_k^{(1)} - y_k^{(2)}, y_{k+1}^{(1)} - y_{k+1}^{(2)}, \dots, y_{k+M-1}^{(1)} - y_{k+M-1}^{(2)})^T$, $\boldsymbol{\mu}$ is the mean of the random vector $\mathbf{n} = (n_k, n_{k+1}, \dots, n_{k+M-1})^T$ given $\mathbf{x}^{(1)}$, and $\boldsymbol{\Sigma}$ is the covariance matrix of \mathbf{n} given $\mathbf{x}^{(1)}$. In order to distinguish the noise mean and variance of the order-1 channel approximation from those of the order-2 channel approximation, we add the subscript 1 to them. The mean and the variance can be expressed as:

$$\boldsymbol{\mu}_1 = \mathbf{H} \cdot \mathbf{x}^{(1)} - \mathbf{G} \cdot \mathbf{x}^{(1)} + \mathbf{S}' \cdot \mathbf{D}^{(1)} \cdot \boldsymbol{\delta}^{(1)}, \quad (9)$$

$$\boldsymbol{\Sigma}_1 = \sigma_J^2 \mathbf{S}' \cdot (\mathbf{D}^{(1)})^2 \cdot \mathbf{S}'^T + \sigma_W^2 \mathbf{F} \mathbf{F}^T. \quad (10)$$

Here, \mathbf{H} , \mathbf{G} , \mathbf{S}' and \mathbf{F} are Toeplitz matrices formed by the sequences $\{\tilde{h}_i\}$, $\{g_i\}$, $\{\tilde{s}'_i\}$ and the equalizer coefficient sequence. For example, if $\tilde{h}_i = 0$ for $i > B$ or $i < -A$, \mathbf{H} is an $M \times (M + A + B)$ matrix with each row equal to a shifted version of the sequence \tilde{h}_i , written in reverse order:

$$\mathbf{H} = \begin{pmatrix} \tilde{h}_B & \tilde{h}_{B-1} & \dots & \tilde{h}_{-A} & 0 & \dots & 0 \\ 0 & \tilde{h}_B & \tilde{h}_{B-1} & \dots & \tilde{h}_{-A} & 0 & \dots \\ \vdots & \vdots & \dots & \dots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & \tilde{h}_B & \tilde{h}_{B-1} & \dots & \tilde{h}_{-A} \end{pmatrix} \quad (11)$$

The matrices \mathbf{G} , \mathbf{S}' and \mathbf{F} are constructed similarly. Since the sequences $\{\tilde{h}_i\}$ and $\{g_i\}$ are generally of different length, the data vectors multiplied by \mathbf{H} and \mathbf{G} will also generally have different lengths. For example, the data bits involved in multiplication with \mathbf{H} are from time index $k - B$ to $k + M - 1 + A$, while the data bits involved in multiplication with \mathbf{G}

are from time index $k - J$ to $k + M - 1$. To simplify the notation, we omit all the indices of the data bits in equations (9) and (10).

$\mathbf{D}^{(1)}$ is a diagonal matrix whose diagonal elements are the transition values d_i . We use the superscript (1) to emphasize that here the transitions are for the recorded sequence $\mathbf{x}^{(1)}$. The column vector $\boldsymbol{\delta}^{(1)}$ contains the net transition shifts for each transition d_i , given that $\mathbf{x}^{(1)}$ is recorded. Similarly, the size of $\mathbf{D}^{(1)}$ and the length of $\boldsymbol{\delta}^{(1)}$ are determined by the range of \tilde{s}' .

We can see from equations (9) and (10), that both the mean and the covariance matrix depend on the transmitted data. However, only the mean is affected by NLTS terms.

2) *Order-2 Channel Approximation*: In the order-2 channel approximation, the noise at time i can be written as

$$n_i = \sum_j x_{i-j}^{(1)} \tilde{h}_j - \sum_j x_{i-j}^{(1)} g_j + \sum_j d_{i-j}^{(1)} (\delta_{i-j} + a_{i-j}) \tilde{s}'_j + \sum_j d_{i-j}^{(1)} \frac{(\delta_{i-j} + a_{i-j})^2}{2} \tilde{s}''_j + w_i, \quad (12)$$

where $\{\tilde{s}''_i\}$ is the sequence of equalized second derivative samples of the transition response.

Because of the second derivative term in equation (12), the noise is no longer Gaussian in nature. The joint distribution of the noise is complicated and the exact pairwise error probability cannot be calculated easily.

Therefore, we approximated the pairwise error probability using the Q -function as in equation (8). Of course, the mean and the covariance matrix of the noise are different from those corresponding to the order-1 channel approximation. They have the following form

$$\boldsymbol{\mu}_2 = \mathbf{H} \cdot \mathbf{x}^{(1)} - \mathbf{G} \cdot \mathbf{x}^{(1)} + \mathbf{S}' \mathbf{D}^{(1)} \boldsymbol{\delta}^{(1)} + \frac{\sigma_j^2}{2} \mathbf{S}'' \mathbf{D}^{(1)} \mathbf{1} + \mathbf{S}'' \mathbf{D}^{(1)} \mathbf{Q}^{(1)} \boldsymbol{\delta}^{(1)} / 2 \quad (13)$$

$$\boldsymbol{\Sigma}_2 = \sigma_j^2 (\mathbf{S}' + \mathbf{S}'' \mathbf{Q}^{(1)}) (\mathbf{D}^{(1)})^2 (\mathbf{S}' + \mathbf{S}'' \mathbf{Q}^{(1)})^T + \sigma_j^4 \mathbf{S}'' (\mathbf{D}^{(1)})^2 \mathbf{S}''^T / 2 + \sigma_W^2 \mathbf{F} \mathbf{F}^T. \quad (14)$$

Here, \mathbf{H} , \mathbf{G} , \mathbf{S}' , $\mathbf{D}^{(1)}$ and \mathbf{F} are the same as in equations (9) and (10). Matrix \mathbf{S}'' comes from the second derivative term. It is a Toeplitz matrix in which each row is a shifted version of the sequence $\{\tilde{s}''_i\}$, in reverse order. The matrix \mathbf{S}'' can be made to have the same size as \mathbf{H} and \mathbf{S}' since we can always find A and B such that $\tilde{h}_j = 0$, $\tilde{s}'_j = 0$ and $\tilde{s}''_j = 0$ when $i < -A$, or $i > B$. The vector $\mathbf{1}$ in (13) represents the all-ones column vector.

Matrix $\mathbf{Q}^{(1)}$ in equations (13) and (14) is a diagonal matrix whose diagonal elements are the net transition shifts δ_i for the recorded sequence $\mathbf{x}^{(1)}$.

Comparing the noise mean and covariance matrix in the order-1 and order-2 channel approximations, we see that the NLTS affects only the noise mean in the order-1 channel approximation, while in the order-2 channel approximation, NLTS affects both the noise mean and the covariance matrix. Similarly, the jitter noise variance appears only in the covariance matrix calculation in the order-1 channel approximation while in the order-2 channel approximation, it appears both in

the mean and the covariance matrix calculation. The AWGN noise variance only appears in the covariance matrix calculation for both cases.

B. Upper and Lower Bounds on Bit-Error-Rate

For the Viterbi detector, the union bound on the probability that an error event occurs at time k can be expressed by the summation of the probabilities of all the possible error events that start at time k and end at time $k + M$ for any given recorded data sequence $\mathbf{x}^{(1)}$. If we group the error events according to the error event lengths, the union bound for the sequence error probability can be written as

$$P_E < \sum_{M=M_{\min}}^{\infty} \sum_{\mathbf{x}^{(1)}} \Pr(\mathbf{x}^{(1)}) \sum_{\substack{\mathbf{x}^{(2)} \text{ such that} \\ \epsilon_M = (\xi_1 \rightarrow \xi_2) \in E_M}} \Pr(\epsilon_M | \mathbf{x}^{(1)}) \quad (15)$$

where M_{\min} is the minimum length of all the error events, and E_M is the set of error events of length M .

The union bound on the bit error probability can be derived from P_E . Denote by $N_b(\epsilon_M)$ the number of erroneous bits corresponding to the error event ϵ_M , i.e., the number of bits by which sequence $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ differ. The bit error probability is the probability that the bit belongs to the set of incorrect bits of an error event. The union bound for the bit error probability is thus given by

$$P_b < \sum_{M=M_{\min}}^{\infty} \sum_{\mathbf{x}^{(1)}} \Pr(\mathbf{x}^{(1)}) \sum_{\substack{\mathbf{x}^{(2)} \text{ such that} \\ \epsilon_M = (\xi_1 \rightarrow \xi_2) \in E_M}} \Pr(\epsilon_M | \mathbf{x}^{(1)}) N_b(\epsilon_M) \quad (16)$$

A lower bound on the BER can be obtained by limiting the summation in the union bound to a set of mutually disjoint error events. The collection of minimum-length error events form such a set. In a minimum-length error event, the two paths ξ_1 and ξ_2 diverge at time k and remerge after the shortest possible time. This occurs when $\mathbf{x}^{(2)}$ differs from $\mathbf{x}^{(1)}$ only in the bit at time k . The minimum error event length is therefore determined by the number of trellis states; specifically, $M_{\min} = J + 1$. For a given recorded sequence $\mathbf{x}^{(1)}$, there is only one such error event because the inputs are binary and there is only one sequence $\mathbf{x}^{(2)}$ that can differ from $\mathbf{x}^{(1)}$ only at time k . Therefore, the minimum-length error events are pairwise disjoint since they correspond to different recorded sequences. It follows that, for each minimum-length error event $\epsilon_{M_{\min}}$, the number of erroneous bits is $N_b(\epsilon_{M_{\min}}) = 1$. In the scenarios we considered, the BER is dominated by single-bit errors [4] corresponding to minimum-length error events. We therefore expect the lower bound based upon such events to be fairly good.

Assuming that the recorded data sequences are equiprobable, the lower bound on the bit error probability obtained by minimum-length error events can thus be written as follows:

$$P_b > \frac{1}{2^{L_E}} \sum_{\mathbf{x}^{(1)}} \Pr(\epsilon_{M_{\min}} = (\xi_1 \rightarrow \xi_2) | \mathbf{x}^{(1)}) \quad (17)$$

TABLE I
NORMALIZED NLTS FOR FOUR TRANSITION PATTERNS

Transition patterns	...001 (1)	...011 (1)	...010 (1)	...000 (1)
NLTS / B	20%	12%	8%	0

where L_E is the effective calculation length of the recorded sequence. The effective calculation length is the length of the span within the recorded sequence that figures in the computation of the pair-wise error probability. For example, suppose we have two sequences that agree in positions $k - T_1$ to $k + T_2$. If the pair-wise error probabilities corresponding to a single-bit error at time k with respect to these two sequences differ, then we would need to use an effective calculation length $L_E > T_2 - T_1 + 1$.

Clearly, the value of L_E is determined by the ISI channel memory, the memory of the noise correlation, the data-dependent noise memory, and the memory of the NLTS. In our model of NLTS, the data bits are written sequentially and the net transition shift of the current transition is affected by previous transition positions, which in turn have been affected by the positions of transitions preceding them. Thus, the NLTS may have, in effect, unbounded memory. This implies that L_E would have to be very large, making the computation of the lower bound impractical. Consequently, in our calculations we use a smaller value for L_E , resulting in only an estimate of the lower bound.

By considering only single-bit error events in the lower bound, we can take advantage of other computational simplifications. For example, the vector λ in this situation is $\lambda = 2x_k^{(1)}(g_0, g_1, \dots, g_J)^T$. Since $x_k^{(1)}$ is either $+1$ or -1 , we can evaluate in advance vectors such as $|\lambda|^T \cdot \mathbf{H}$ that, when suitably truncated, are used in the calculation of the pairwise error probabilities.

IV. SIMULATION RESULTS

To compare the BER estimates derived in the previous section to the Monte-Carlo simulation results presented in [4], we calculated the estimate of the lower bound for both the order-1 and order-2 channel approximations, using the same channel parameters. We use a minimum mean-squared error (MMSE) equalizer design with monic constraint [13]. The equalizer is a 21-tap FIR filter and the equalization target has 3 taps. The equalizer output serves as the input to a Viterbi detector matched to the target.

For the NLTS calculation, we set the medium to soft-underlayer spacing to 20nm, and the medium thickness is set to 10nm. The channel bit spacing is 16nm, corresponding to a linear density of about 1.59×10^6 bits/inch. The remanent magnetization to head field gradient ratio is set to 1.5. With these parameters, the NLTS of the isolated dibit pattern is about 20% (absolute value) of the channel bit spacing. In Table I, we list the absolute value of the normalized NLTS values for several input patterns.

In Fig. 2, the BER lower bound calculated by equation (17) is shown for the dibit precompensation scheme using the order-1 channel approximation. The x -axis represents the

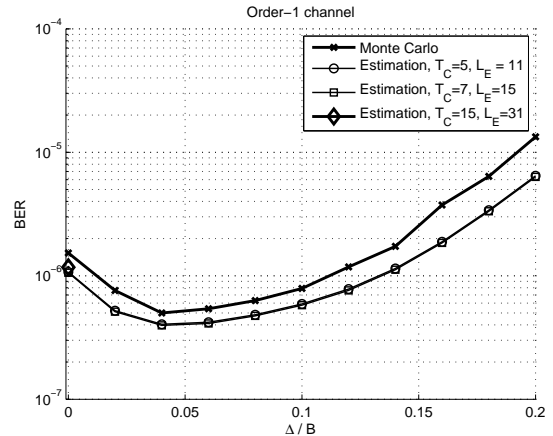


Fig. 2. The lower bound estimate for order-1 channel approximation with dibit precompensation scheme. $T_{50}/B = 1$; $\sigma_J/B = 0.08$; $SNR_W = 29$ dB; $B = 16$ nm. GPR target = [1,0.76,-0.06].

precompensation value normalized by channel bit spacing. The data sequence $\mathbf{x}^{(1)}$ we considered in the calculation is from time $k - T_C$ to $k + T_C$, where T_C is shown in the legend of Fig. 2. Therefore, the effective calculation length that we use is given by $L_E = 2T_C + 1$. We can see that the curves for $L_E = 11$ and $L_E = 15$ are almost identical. For $L_E = 31$, because of the computational complexity, we calculated only one point where no precompensation was used. The result is very close to the results corresponding to $L_E = 11$ and $L_E = 15$.

The BER generated by Monte-Carlo simulation is also shown in the figure. We can see that the lower bound curve for the order-1 channel approximation is very close to the Monte-Carlo simulation result. The optimal precompensation value that minimizes the BER can also be deduced from the lower bound estimate.

Fig. 3 shows the estimate of the lower bound for the order-2 channel approximation for dibit precompensation. The channel parameters are the same as in Fig. 2. The Monte-Carlo simulation results are also shown.

We can see that the lower bound is not as tight as for the order-1 channel approximation because of the inaccuracy introduced by the Gaussianity assumption in calculating the pairwise error probability. However, the optimal precompensation value obtained by using the estimate is the same as the one obtained by means of Monte-Carlo simulation. (We note that, again, the difference between results obtained by setting $L_E = 11$ and $L_E = 15$ is very small.) Moreover, the computation time required for the lower bound estimate was much less than that required for the Monte-Carlo simulation. For example, the curve in Fig. 3 corresponding to the Monte-Carlo simulation required several hours to compute, whereas the calculation of the lower bound estimate curve in the same figure took only several minutes.

Table II shows the optimal precompensation values for the two-level precompensation and three-level (2-bit look-back) precompensation schemes. As in the case of dibit precompensation, the optimal precompensation values derived from the lower bound estimate are very close to those obtained

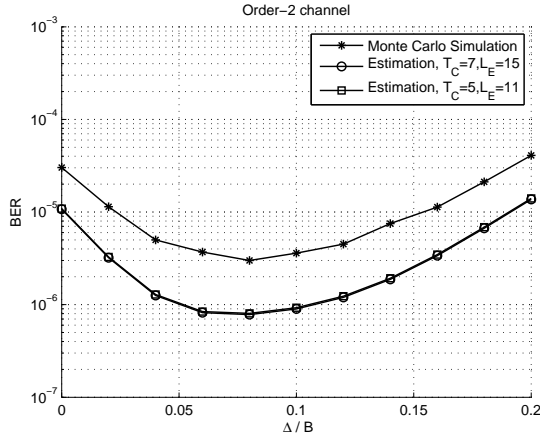


Fig. 3. The lower bound estimate for the order-2 channel approximation with a dibit precompensation scheme. $T_{50}/B = 1$; $\sigma_J/B = 0.08$; $SNR_W = 29\text{dB}$; $B = 16\text{nm}$. GPR target $= [1, 0.76, -0.06]$.

TABLE II
COMPARISON OF THE OPTIMAL PRECOMPENSATION VALUES OBTAINED FROM THE MONTE-CARLO SIMULATION AND THE LOWER BOUND ESTIMATE FOR THE ORDER-2 CHANNEL APPROXIMATION.

$T_{50}/B = 1$; $\sigma_J/B = 0.08$; $SNR_W = 29\text{dB}$; $B = 16\text{nm}$. GPR target $= [1, 0.76, -0.06]$.

	Dibit pre-comp	Two-level Precomp	2-bit look-back Precomp
Monte-Carlo Simulation	0.08	(0.09, 0.05)	(0.12, 0.06, 0.1)
Lower Bound Estimation	0.08	(0.088, 0.043)	(0.114, 0.056, 0.093)

by Monte-Carlo simulation, while requiring substantially less time to generate.

For the two-level precompensation scheme, the table specifies $(\Delta_H/B, \Delta_L/B)$, where Δ_H and Δ_L are the precompensation values for the rightmost transition in the recorded transition patterns “011” and “111”, respectively. For the 2-bit look-back precompensation scheme, the precompensation values $(\Delta_H, \Delta_M, \Delta_L)$ correspond to transition patterns (“011”, “111”, “101”). In the Monte-Carlo simulation, we used a step size of 0.01 when optimizing the precompensation values for the two-level precompensation scheme, and 0.02 for the other schemes. We can see that the values obtained from the lower bound estimate are fairly close to those found by simulation.

Except in the case of the dibit precompensation scheme, we used the MATLAB optimization solver `fminunc` to compute the optimal precompensation values in the analysis. This optimization routine determines the minimum value of a possibly multivariate function, under the assumption that its scalar or vector input is unconstrained. In our application, the function is the BER estimate, and the inputs are the precompensation values of the specified precompensation scheme.

V. FURTHER APPLICATIONS OF THE BER ANALYSIS

The BER analysis sheds light on some of the simulation results presented in [4]. In that paper, we observed that in some cases, the two-level precompensation scheme can

perform better than the system without NLTS and with no write precompensation. A comparison of the error events seen in the system with optimized two-level precompensation to those seen in the no-NLTS case showed that the single-bit error events were dominant in both settings, but with fewer occurrences in the precompensated system.

Table III provides a more detailed comparison of the frequency of occurrence of single-bit error events in the two systems as observed in Monte-Carlo simulation. The error events in the table are denoted as follows. A symbol “1” (resp. “0”) denotes a position where the recorded sequence has the input value +1 (resp. -1). The symbol - indicates an error in which a +1 was detected as a -1, while a + indicates an error in which a -1 became a +1. Thus the patterns represent a span of five bits centered around the error position. For example, error event 0 0 - 0 0 means that the recorded data pattern was 0 0 1 0 0 and the detected one was 0 0 0 0 0.

We can see that two-level precompensation reduces the number of occurrences for some error events while increasing it for others. To see the implications of this, consider the length-6 data sequence 0 0 1 0 0 0. The table shows that within the first five bits, the system with two-level precompensation would experience fewer error events of type 0 0 - 0 0 than the no-NLTS system. On the other hand, within the last five bits, the precompensated system would suffer more single-bit error events of type 0 1 + 0 0. A possible explanation for this is the shift of the second transition (between the 1 and the 0) induced by the NLTS and two-level precompensation. Overall, we see that the total number of error counts for the no-NLTS case is 957, higher than the 827 found in the system using two-level precompensation. (It should be noted that, in addition to the single-bit error events reflected in the table, both systems suffered a small number of multi-bit error events.)

The differences in the distribution of the single-bit error events for the two systems can be visualized in another way, as shown in Fig. 4. Since the minimum-length error events correspond to single-bit error events, we examined for both systems the histogram of the pairwise error probabilities for all the minimum-length error events observed when $L_E = 11$. The x -axis in the plot represents the logarithm of the pairwise error probability. It is divided into bins of size 0.1 and the histogram shows the number of error events with probability falling in each bin. The BER is dominated by those error events that have large pairwise error probability, highlighted in the figure. We can see that the error events that have the largest pairwise error probability in the no-NLTS case are concentrated in the rightmost peak of the distribution. In contrast, in the system with two-level precompensation, the error events span a wider range along the x -axis, with many fewer events falling in each bin. There are a small number of error events whose pair-wise error probability increases very slightly, but a large number of error events that have a decreased pair-wise error probability. The number of error events with the highest error probability is much less than that found in the no-NLTS case. The overall BER decreases from 3.59×10^{-7} to 3.03×10^{-7} , according to the lower bound estimate.

The change in the pairwise error probability comes from

TABLE III
MONTE-CARLO SIMULATION: SINGLE-ERROR EVENT COUNTS FOR SYSTEM WITH NO NLTS AND WITH OPTIMAL TWO-LEVEL PRECOMPENSATION

Two-level Precomp. Improves			Two-level Precomp. Degrades		
Event	No NLTS	2-level	Event	No NLTS	2-level
00-00	118	50	00-10	1	7
00-01	108	60	00-11	0	8
01+10	116	110	01+00	1	34
01+11	117	93	01+01	0	33
10-00	109	100	10-01	98	103
11+10	123	76	10-10	0	21
11+11	93	46	10-11	0	30
10+10	2	0	11+00	0	8
10+11	2	0	11+01	0	5
Total	788	535	Total	100	249

* Total number of error counts for no-NLTS is 957. Total number of error counts for two-level precomp. is 827. 5.12×10^8 bits simulated. Same condition as Fig.3.

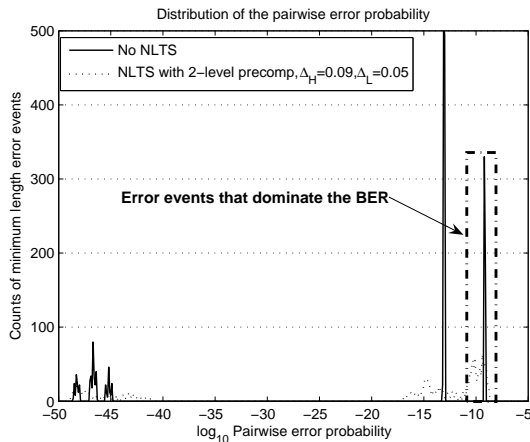


Fig. 4. The histogram of the pairwise error probability for a system with no NLTS and a system with optimal two-level precompensation. $T_{50}/B = 1$; $\sigma_J/B = 0.08$; $SNR_W = 29$ dB; $B = 16$ nm. GPR target=[1,0.76, 0.06].

the change in the effective distance between the correct path and the incorrect path in the two systems. The histogram shows that for many single-bit error events, especially the dominant ones, the effective distance is larger in the two-level precompensation system than in the no-NLTS case. Thus, the BER is improved.

We can also apply our BER analysis as a tool in the optimization of precompensation schemes with a larger number of levels, for which a brute force search based upon Monte-Carlo simulation is not feasible. This permits a more extensive comparison of the BER performance of multilevel precompensation schemes.

For example, in Fig. 5, we compare the results of Monte-Carlo simulation performance for several precompensation schemes using the order-2 channel approximation. The channel parameters are the same as in Fig. 3. The performance curves for the dibit and two-level precompensation schemes are taken from [4]. The figure also includes the performance curve for the 2-bit look-back scheme and the 3-bit look-back

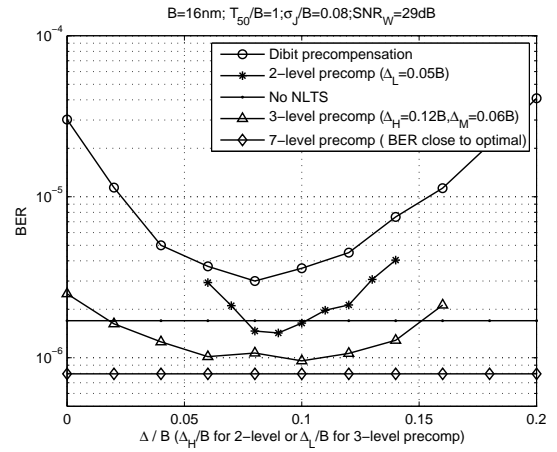


Fig. 5. Comparison of the dibit, two-level and multilevel precompensation schemes for $B=16$ nm, using an order-2 channel approximation. All the parameters are the same as in Fig. 3.

scheme. For the 2-bit look-back scheme, we first optimized the precompensation values using the BER estimate. With the resulting values as a starting point, we then ran Monte-Carlo channel simulations using points nearby to generate the curve. For the 3-bit look-back scheme, we show only one BER simulation result, corresponding to the optimal precompensation values obtained from the analytic BER estimate. We see that the 2-bit and the 3-bit look-back schemes can provide a slight improvement in the BER, relative to the two-level precompensation scheme. However, the improvement is not as significant as that found in going from dibit precompensation to two-level precompensation.

Simulation results corresponding to a higher user density are presented in Fig. 6. Here we fix the head media parameters the same as those used to generate Fig. 5 and only push the linear density by setting the channel bit spacing B to 14nm. With the increased channel density, the normalized NLTS of the second transition in an isolated dibit pattern is increased from 20% to approximately 26%. The normalized channel density T_{50}/B is about 1.14 and σ_J/B is about 0.09. The SNR_W is assumed to be 29dB. As in Fig. 5, performance curves for dibit, two-level, 2-bit look-back and 3-bit look-back precompensation schemes are shown.

By pushing the linear density, NLTS increases and the precompensation values for all the levels are generally larger than those in Fig. 5. We see that the two-level precompensation scheme no longer gives better performance than a system without NLTS in this situation, whereas the 2-bit and 3-bit look-back schemes both do. However, we also note that the 7-level 3-bit look-back scheme offers only a slight advantage over the 3-level 2-bit look-back scheme. This may indicate that just adding more levels of precompensation may not be enough to overcome the deleterious effects of increased ISI and higher normalized jitter noise when pushing linear densities higher.

VI. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we studied the performance of perpendicular recording systems with AWGN, transition jitter noise, and

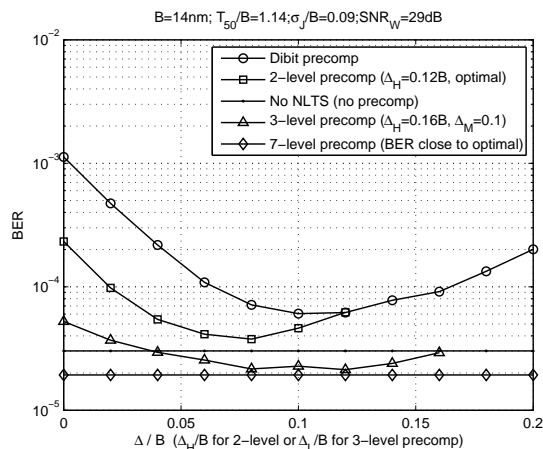


Fig. 6. Comparison of the Dibit, Two-level and multilevel precompensation schemes for $B=14\text{nm}$, using order-2 channel approximation. All other parameters are the same as in Fig. 5. Therefore, $T_{50}/B \approx 1.14$; $\sigma_J/B \approx 0.09$; dibit NLTS $\approx 26\%$. GPR target $= [1, 0.84, 0.05]$.

nonlinear transition shift (NLTS). We assumed the use of a Viterbi detector, and presented an analytic lower bound on the resulting BER. For an order-1 channel approximation, an easily computed estimate of the lower bound was shown to give BER results very close to the actual lower bound. For both order-1 and order-2 channel approximations, the optimal precompensation values obtained by using the lower bound estimate were found to be very close to those obtained from Monte-Carlo performance simulation.

We also used the lower bound estimate to study the potential performance gain offered by multilevel precompensation schemes with respect to the simpler dibit and two-level precompensation schemes. Our results show that although 2-bit and 3-bit look-back schemes provide some reduction in BER, the improvement is modest, especially for higher linear densities where the effects of increased ISI and jitter noise are more difficult to combat.

There are a number of topics related to this study that deserve future consideration. A thorough comparison of different design criteria for write precompensation schemes is needed. We have already accomplished several comparisons. We compared two criteria – minimizing the mean-square-error (MSE) between the channel output signal and the output of an ideal noiseless channel, and minimizing the variance of the net transition shifts – to the min-BER criterion in [4]. Here, we briefly present the comparison results of a third criterion to the min-BER criterion: optimizing write precompensation values based upon the characterization of NLTS by the well-known dipulse extraction method [14]. By extracting the dipulse from the waveform of pseudo-random sequences, the “echoes” in the extracted dipulse response corresponds to the NLTS. Therefore, the precompensation values can be tuned to minimize these “echoes”. We searched for the optimal dibit and two-level precompensation values that minimize the echoes, in other words, minimize the squared error of the extracted dipulse response at the echoes region. Under the same conditions as in Fig. 3, the optimal dibit precompensation value found by both approaches was

$0.08B$. For two-level precompensation, the best precompensation values according to the echo-minimization criterion are $(\Delta_H, \Delta_L) = (0.1B, 0.07B)$, very close to the values obtained by minimizing BER.

A more interesting problem is to extend the analysis in this paper to systems using more sophisticated detectors, such as NPML and PDNP detectors. There have been several analytical studies of PDNP detector performance in the absence of NLTS; see, for example [11], [12]. A simulation-based comparison of PDNP and Viterbi detector performance in the presence of NLTS can be found in [15]. The development of analytical tools to help in the design and optimization of write precompensation schemes for use with these advanced detectors remains an area for future research.

ACKNOWLEDGEMENTS

The authors would like to thank the Information Storage Industry Consortium (INSIC) Extremely High Density Recording (EHDR) program and the Center for Magnetic Recording Research at UC San Diego for funding this work. They also wish to thank the anonymous reviewers, whose insightful comments and valuable suggestions helped to improve the paper.

REFERENCES

- [1] X. Che, “Nonlinearity measurements and write precompensation studies for a PRML recording channel,” *IEEE Trans. Magn.*, vol. 31, no. 6, pp. 3021–3026, Nov. 1995.
- [2] A. Taratorin, J. Fitzpatrick, S. X. Wang, and B. Wilson, “Non-linear interactions in a series of transitions,” *IEEE Trans. Magn.*, vol. 33, no. 1, pp. 956–961, Jan. 1997.
- [3] F. Lim and A. Kavčić, “Optimal precompensation for partial erasure and nonlinear transition shift in magnetic recording using dynamic programming,” in *Proc. IEEE Global Telecommun. Conf., Globecom’05*, vol. 1, St. Louis, Missouri, USA, Nov. 28–Dec. 2 2005, pp. 58–62.
- [4] Z. Wu, H. N. Bertram, P. H. Siegel, and J. K. Wolf, “Nonlinear transition shift and write precompensation in perpendicular magnetic recording,” in *Proc. IEEE Conf. Commun., ICC’08*, Beijing, China, May 19–23 2008, pp. 1972–1976.
- [5] Z. Wu, “Channel modeling, signal processing and coding for perpendicular magnetic recording,” Ph.D. dissertation, Department of Electrical and Computer Engineering, University of California, San Diego, USA, 2009.
- [6] H. N. Bertram, *Theory of Magnetic Recording*. Cambridge University Press, 1994.
- [7] K. Nakamoto and H. N. Bertram, “Analytic perpendicular-recording model for transition parameter and NLTS,” *The Magnetic Society of Japan*, vol. 26, no. 2, pp. 79–85, 2002.
- [8] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York, NY, USA: McGraw-Hill, Inc., 1979.
- [9] L. C. Barbosa, “Maximum likelihood sequence estimators: A geometric view,” *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 419–427, Mar. 1989.
- [10] R. He and N. Nazari, “An analytical approach for performance evaluation of partial response systems in the presence of signal-dependent medium noise,” in *Global Telecommunications Conference, 1999. GLOBECOM ’99*, vol. 1, Rio de Janeiro, Dec. 1999, pp. 939–943.
- [11] A. Kavčić and J. M. Moura, “The Viterbi algorithm and Markov noise memory,” *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 291–301, Jan. 2000.
- [12] J. Moon and J. Park, “Pattern-dependent noise prediction in signal-dependent noise,” *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 730–743, Apr. 2001.
- [13] J. Moon and W. Zeng, “Equalization for maximum likelihood detectors,” *IEEE Trans. Magn.*, vol. 31, no. 2, pp. 1083–1088, Mar. 1995.
- [14] D. Palmer, P. Ziperovich, R. Wood, and T. Howell, “Identification of nonlinear write effects using pseudorandom sequences,” *Magnetics, IEEE Transactions on*, vol. 23, no. 5, pp. 2377–2379, Sep 1987.

- [15] Z. Wu, P. H. Siegel, J. K. Wolf, and H. N. Bertram, "Mean-adjusted pattern dependent noise prediction for perpendicular recording channels with nonlinear transition shift," *IEEE Trans. Magn.*, vol. 44, no. 11, pp. 3761–3764, Nov. 2008.

PLACE
PHOTO
HERE

Zheng Wu received her Bachelor's and Master's degree in Tsinghua University, Beijing, China in 2000 and 2003, and the Ph.D degree in University of California at San Diego, USA in 2009. Currently, she is a staff engineer at Link_A_Media Devices Corp., Santa Clara, California, USA. Her research interests include signal processing and coding techniques in magnetic recordings and storage.



Paul H. Siegel (M'82-SM'90-F'97) received the S.B. and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1975 and 1979, respectively.

He held a Chaim Weizmann Postdoctoral Fellowship at the Courant Institute, New York University. He was with the IBM Research Division in San Jose, CA, from 1980 to 1995. He joined the faculty of the School of Engineering at the University of California, San Diego in July 1995, where he is currently Professor of Electrical and Computer Engineering.

He is affiliated with the California Institute of Telecommunications and Information Technology, the Center for Wireless Communications, and the Center for Magnetic Recording Research where he holds an endowed chair and currently serves as Director. His primary research interests lie in the areas of information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission.

Prof. Siegel was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and was re-elected for a 3-year term in 2009. He served as Co-Guest Editor of the May 1991 Special Issue on "Coding for Storage Devices" of the *IEEE TRANSACTIONS ON INFORMATION THEORY*. He served the same *TRANSACTIONS* as Associate Editor for Coding Techniques from 1992 to 1995, and as Editor-in-Chief from July 2001 to July 2004. He was also Co-Guest Editor of the May/September 2001 two-part issue on "The Turbo Principle: From Theory to Practice" of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*. He was co-recipient, with R. Karabed, of the 1992 IEEE Information Theory Society Paper Award and shared the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award with B. Marcus and J.K. Wolf. With J.B. Soriaga and H.D. Pfister, he received the 2007 Best Paper Award in Signal Processing and Coding for Data Storage from the Data Storage Technical Committee of the IEEE Communications Society. He holds several patents in the area of coding and detection, and was named a Master Inventor at IBM Research in 1994. He is a member of Phi Beta Kappa and the National Academy of Engineering.



Jack Keil Wolf received the B.S.E.E. degree from the University of Pennsylvania Philadelphia, in 1956, and the M.S.E., M.A., and Ph.D. degrees from Princeton University, Princeton, NJ, in 1957, 1958, and 1960, respectively. He is currently the Stephen O. Rice Professor of Electrical and Computer Engineering and a member of the Center for Magnetic Recording Research at the University of California-San Diego, La Jolla. He is also a Vice President, Technology (part time) at Qualcomm, Inc., San Diego.

His current interest is in signal processing for storage systems.