Optimizing the Bit Aspect Ratio of a Recording System Using an Information-Theoretic Criterion

Shaohua Yang, *Student Member, IEEE*, Aleksandar Kavčić, *Member, IEEE*, and William Ryan, *Senior Member, IEEE*

***Abstract*— A method is developed to determine the optimal bit aspect ratio of a recording system which maximizes the information density per unit area of the recording medium. The areal information density is computed as the achievable mutual information rate divided by the area of each binary symbol on the medium. A data-dependent autoregressive noise model is fitted to the recording channel and the achievable information rate of the recording channel is estimated by computing the achievable information rate of the channel model.**

***Index Terms*—Areal density, bit aspect ratio, channel capacity, channel modeling, information rate, magnetic recording channel, microtrack model.**

I. Introduction

THE BIT aspect ratio of a recording system is the ratio between the symbol separation T (in nanometers) and the track width W (in nanometers). The parameters T and W, together with the head-disk interface and tracking mechanism, affect the intersymbol interference (ISI) and signal-to-noise ratio (SNR) of the data channel which in turn affect the channel capacity.

We define the areal information density ρ as the number of information bits that can be reliably written per unit area of the recording medium surface. We measure information by the achievable mutual information rate $\mathcal{I}(X;Y)$ between the channel input and output. Thus, we define

$$\rho \stackrel{\triangle}{=} \frac{\mathcal{I}(X;Y)}{TW}. \tag{1}$$

We assume throughout that the head-disk interface and tracking mechanism are fixed. Thus, the achievable mutual information rate $\mathcal{I}(X;Y)$ and the areal density ρ are affected only by the parameters T and W, i.e., $\rho = \rho(T,W)$. By selecting optimal values of parameters T and W, we may maximize the value of the areal information density $\rho(T,W)$.

In [1], a method to optimize the areal parameters T and W was proposed by analyzing the bit-error rate performance of partial response maximum likelihood (PRML) recording channels. Earlier references using the same criterion can be found in [1].

In [2], the symbol separation T was optimized to maximize the linear information density (in bits/in) along a track

$$T^* = \arg\max_{T} \frac{\mathcal{I}(X;Y)}{T}. \tag{2}$$

In this paper, using the same information-theoretic criterion as in [2], we develop a method to optimize both of the parameters T and W of a recording system by maximizing the areal information density $\rho(T,W)$

$$(T^*, W^*) = \arg\max_{\{T,W\}} \frac{\mathcal{I}(X;Y)}{TW} \tag{3}$$

where the values of T and W also depend on limitations set by other system specifications.

For a real recording channel, which is characterized by complicated signal nonlinearities and data-dependent noise correlation, the computation of the achievable information rate $\mathcal{I}(X;Y)$ is still an open problem. Here, we apply the method in [3] to estimate $\mathcal{I}(X;Y)$. We first fit a data-dependent autoregressive (AR) noise model [4] to the recording channel, which captures the first- and second-order statistics of the channel output waveform. We then estimate the achievable information rate of the channel by computing the information rate of the model. The achievable information rate of a data-dependent AR noise model can be accurately computed by the Monte Carlo method proposed in [5] and [6].

II. Data-Dependent Autoregressive Noise Model

The data-dependent AR noise model is essentially the same as in [4]. We briefly restate it to introduce the notation. Let the binary channel input sequence be X_k and its realization be x_k. Let the channel output sequence be Y_k, and its realization be y_k. The channel noise N_k is additive and is assumed to be generated by a data-dependent AR filter driven by an independent and identically distributed (i.i.d.) zero-mean unit-variance Gaussian noise sequence W_k. The realization of random variables N_k and W_k are denoted by n_k and w_k, respectively. We define the data-dependent AR noise channel model as

$$y_k = g\left(x_{k-I}^{k+I}\right) + n_k \tag{4}$$

$$n_k = \sum_{i=1}^{L} b_i\left(x_{k-D}^{k+D}\right) n_{k-i} + \sigma\left(x_{k-D}^{k+D}\right) w_k. \tag{5}$$

Here, I, D, and L are nonnegative integers, $g\left(x_{k-I}^{k+I}\right)$ is the noiseless signal, which depends on the $2I+1$ channel

Manuscript received January 6, 2003. This work was supported by the Information Storage Industry Consortium.

S. Yang and A. Kavčić are with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA.

W. Ryan is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 98571 USA.

Digital Object Identifier 10.1109/TMAG.2003.815437

inputs x_{k-I}^{k+I}, and the AR filter tap-coefficients $b_i\left(x_{k-D}^{k+D}\right)$ and $\sigma\left(x_{k-D}^{k+D}\right)$ are functions of the $2D+1$ channel inputs x_{k-D}^{k+D}. The values $2I+1$ and $2D+1$ are the data-dependent window sizes of the signal and the AR filter, respectively, and L is the Markov memory length of the noise. The parameters (I, D, L) specify the *size* of the model. By properly selecting the parameters (I, D, L), this channel model can accurately model the data-dependence of the signal and capture the first- and second-order statistics of the channel noise.

The data-dependent AR noise model is an indecomposable finite-state machine channel. If we let integers M_1 and M_2 be

$$M_1 = \max(I, D) \tag{6}$$

$$M_2 = \min(-I-L, -D) \tag{7}$$

then the total channel memory length of this model is $M = M_1 - M_2 + 1$. We denote the state of the channel by S_k. Assuming that i is the value of the integer defined by the binary input string $x_{k+M_2+1}^{k+M_1}$, we let $s_k = i$ be the state realization at time k. Here, s_k satisfies $0 \le s_k \le 2^{M-1} - 1$. Given the initial state s_0 of the channel, the channel state sequence s_k and the channel input sequence x_k uniquely determine each other. Hence, in the following, we interchangeably use the state symbol S for the input symbol X. With such a definition of the channel state, we have the conditional probability density function of the channel output

$$f_{Y_k|Y_{-\infty}^{k-1}, S_{-\infty}^{k}}\left(y_k|y_{-\infty}^{k-1}, s_{-\infty}^{k}\right) = f_{Y_k|Y_{k-L}^{k-1}, S_{k-1}^{k}}\left(y_k|y_{k-L}^{k-1}, s_{k-1}^{k}\right). \tag{8}$$

As shown in [4], given a fixed model size (I, D, L), we can estimate the model parameters from a sufficiently long channel input sequence $x_{1+M_2}^{K+M_1}$ and its associated channel output sequence y_1^K, e.g., $K > 10^6$. For each possible channel input pattern $x_{k-I}^{k+I} = \alpha$, we estimate the noiseless signal lookup value $g(\alpha)$ by

$$g(\alpha) \approx \frac{1}{K_\alpha} \sum_{k: x_{k-I}^{k+I} = \alpha} y_k \tag{9}$$

where K_α is the number of the occurrences of the pattern $x_{k-I}^{k+I} = \alpha$ in the channel input sequence $x_{1+M_2}^{K+M_1}$.

Similarly, for each possible input pattern $x_{k-D}^{k+D} = \beta$, we estimate the data-dependent noise covariance $c_{i,j}(\beta) = \mathrm{E}\left[N_{k-i}N_{k-j}\,\middle|\,X_{k-D}^{k+D} = \beta\right]$ as

$$c_{i,j}(\beta) \approx \frac{1}{K_\beta} \sum_{k: x_{k-D}^{k+D} = \beta} \cdot \left(y_{k-i} - g\left(x_{k-i-I}^{k-i+I}\right)\right)\left(y_{k-j} - g\left(x_{k-j-I}^{k-j+I}\right)\right). \tag{10}$$

Here, K_β is the number of occurrences of the pattern $x_{k-D}^{k+D} = \beta$ in the channel input sequence $x_{1+M_2}^{K+M_1}$. We can then compute the lookup values of the AR filter coefficients $b_i(\beta)$ and $\sigma(\beta)$ by solving the following augmented Yule–Walker equations (Wiener–Hopf equations)

$$\mathbf{C}(\beta)\begin{bmatrix} b_1(\beta) \\ b_2(\beta) \\ \vdots \\ b_L(\beta) \end{bmatrix} = \begin{bmatrix} -\sigma^2(\beta) + c_{0,0}(\beta) \\ c_{0,1}(\beta) \\ \vdots \\ c_{0,L}(\beta) \end{bmatrix} \tag{11}$$

where the matrix $\mathbf{C}(\beta)$ is defined as

$$\mathbf{C}(\beta) = \begin{bmatrix} c_{1,0}(\beta) & c_{2,0}(\beta) & \dots & c_{L,0}(\beta) \\ c_{1,1}(\beta) & c_{2,1}(\beta) & \dots & c_{L,1}(\beta) \\ \vdots & \vdots & \ddots & \vdots \\ c_{1,L}(\beta) & c_{2,L}(\beta) & \dots & c_{L,L}(\beta) \end{bmatrix}. \tag{12}$$

III. Estimating the Mutual Information Rate

We estimate the achievable information rate of the channel by computing the information rate of the channel model. For a data-dependent AR noise model, we have

$$\begin{aligned}
\mathcal{I}(X;Y) &= \lim_{n\to\infty} \frac{1}{n} I(X_1^n; Y_1^n\,|S_0) && (13)\\
&\stackrel{(a)}{=} \lim_{n\to\infty} \frac{1}{n} I(S_1^n; Y_1^n\,|S_0) && (14)\\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Big[h(Y_k\,|S_0, Y_1^{k-1}) - h(Y_k\,|S_0^n, Y_1^{k-1})\Big] && (15)\\
&\stackrel{(b)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Big[h(Y_k\,|S_0, Y_1^{k-1}) - h(Y_k\,|S_{k-1}^k, Y_{k-L}^{k-1})\Big] && (16)
\end{aligned}$$

where (a) follows from the fact that the input sequence X_k and the state sequence S_k determine each other uniquely for any given initial state S_0, and (b) is a consequence of (8).

Using the Monte Carlo method in [5], [6], we have

$$\mathcal{I}(X;Y) \approx \frac{1}{n} \sum_{k=1}^{n} \Big[-\log\left(f(y_k\,|s_0, y_1^{k-1})\right) - \frac{1}{2}\log\left(2\pi e \sigma^2\left(x_{k-D}^{k+D}\right)\right)\Big] \tag{17}$$

where n is a large integer. The information rate $\mathcal{I}(X;Y)$ in (17) obviously depends on the channel input distribution. When the channel inputs are independent and uniformly distributed (i.u.d.), the mutual information rate $\mathcal{I}(X;Y)$ is also referred to as the i.u.d. channel capacity C_{iud}, which is the code rate limit of *random* linear error control codes, e.g., turbo codes and low-density parity check (LDPC) codes. For an input distribution optimized by the algorithms in [5]–[7], a higher mutual information rate may be induced [3]. We note that a nonlinear code might be needed to achieve the optimized information rate. In this paper, we only show results for the i.u.d. channel input distribution, so that

$$\rho(T, W) \stackrel{\triangle}{=} \frac{C_{\mathrm{iud}}}{TW}. \tag{18}$$

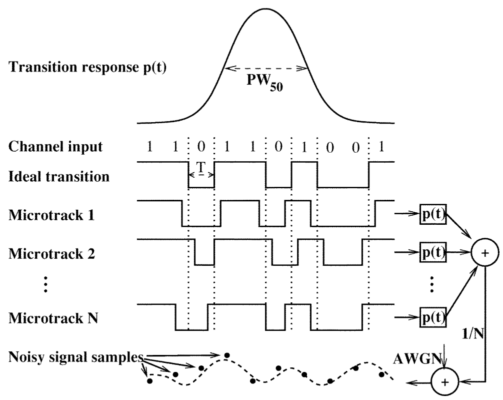

Fig. 1. Microtrack longitudinal recording channel model.

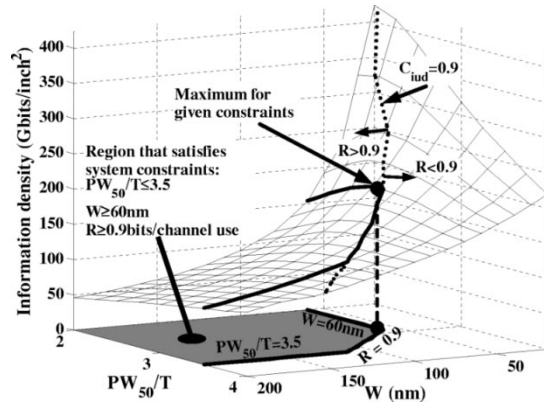

Fig. 2. Maximizing the information density.

The accuracy of the estimated information rate $\mathcal{I}(X;Y)$ is affected by the model size determined by (I, D, L). As the model size is increased, the estimated information rate converges to a single value [3]. We select the model size so that this convergence is achieved.

IV. Application and Results

As an illustration, we optimize the areal parameters T and W of a longitudinal recording system simulated by a microtrack model [8] as depicted in Fig. 1. The recording track is modeled to consist of N parallel microtracks. The width of each microtrack is assumed to be $\mu = 10$ nm (which equals the cross-track correlation width of the magnetic medium). Thus, the width of the recording track is $W = 10 \cdot N$ nm. Each microtrack is one subchannel whose transition response is a Gaussian pulse with a half-amplitude width $PW_{50} = 140$ nm. The media noise is modeled by the transition jitter distortions in each microtrack, i.e., transitions between 1s and 0s drift around their ideal positions by an i.i.d. Gaussian random distance whose mean is zero and variance is $\sigma_j^2 = 537.61$ nm^2. The random jitter processes are assumed to be independent for different microtrack subchannels. Two adjacent transitions in a microtrack cancel each other if the distance between them is less than the percolation threshold $T_{\min} = 31.45$ nm. The channel output is the sum of the subchannel outputs and the additive white Gaussian noise (AWGN). The media noise is defined as the difference between the channel output when N microtracks are read without AWGN and the channel output when $N = \infty$ microtracks are read without AWGN. The one-sided power spectral density N_0 of the AWGN is set so that media noise power is 90% of the total noise power when $N = 10$ and $T = 46.7$ nm. The variance σ^2 of AWGN is linearly proportional to the sampling rate, i.e., $\sigma^2 \propto 1/T$.

We vary the symbol separation T and the number of microtracks N. For each channel configuration (T, W), we estimate the information rate C_{iud} and the information density $\rho(T, W)$ with the help of the data-dependent AR noise model. The numerical results are shown in Fig. 2. As already observed in [2], when the track width W is fixed, the information density $\rho(T, W)$ achieves its maximum at an optimal ratio PW_{50}/T. We notice that the optimal ratio PW_{50}/T increases with W. If we fix the ratio PW_{50}/T, the information density ρ increases when W decreases. However, limitations on track servo accuracy constrain the track width to be greater than some minimum. In general, the optimal system configuration for a recording channel should be selected by constraining the trellis complexity (PW_{50}/T), track width W, and the code rate R to a practically implementable region. As an example, we consider the following system complexity requirement: $PW_{50}/T \leq 3.5$, $W \geq 60$ nm and $R \geq 0.9$ bits/channel-use as depicted in the shaded area in Fig. 2. Under these constraints, ρ achieves its maximum of 200 Gb/in^2 when $PW_{50}/T = 2.8$ and $W = 60$ nm as seen in Fig. 2. Given that $PW_{50} = 140$ nm, we conclude that the optimal symbol dimension for our constraints is $T^* = 50$ nm, $W^* = 60$ nm.

V. Conclusion

With the help of the data-dependent AR noise channel model, we estimate the information rate of a recording system. The areal parameters of a recording system are selected to maximize the information density. This work represents the first instance in which the optimal bit-aspect ratio has been determined using an information-theoretic criterion.

References

[1] W. Huber and G. Worstell, "Maximal areal density for PRML data channels," *IEEE Trans. Magn.*, vol. 32, pp. 3956–3958, Sept. 1996.

[2] W. Ryan, "Optimal code rates for Lorentzian channel models," in *Proc. IEEE GLOBECOM Conf.*.

[3] S. Yang and A. Kavčić, "On the capacity of data-dependent autore-gres-sive noise channels," in *Proc. Allerton Conf. Communications and Control*, 2002.

[4] A. Kavčić and A. Patapoutian, "A signal-dependent autoregressive channel model," *IEEE Trans. Magn.*, vol. 35, pp. 2316–2318, Sept. 1999.

[5] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," *Proc. IEEE ICC 2001*, pp. 2692–2695, June 2001.

[6] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," *Proc. IEEE Globecom 2001*, pp. 2992–2996, Nov. 2001.

[7] A. Kavčić, "On the capacity of Markov sources over noisy channels," *Proc. IEEE GLOBECOM 2001*, pp. 2997–3001, Nov. 2001.

[8] R. D. Barndt, A. J. Armstrong, H. N. Bertram, and J. K. Wolf, "A simple statistical model of partial erasure in thin film disk recording systems," *IEEE Trans. Magn.*, vol. 27, pp. 4978–4980, Nov. 1991.