On the Capacity of Data-Dependent Autoregressive Noise Channels*

Shaohua Yang and Aleksandar Kavčić
Division of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138
yangsh@deas.harvard.edu, kavcic@hrl.harvard.edu

Abstract

Data-dependent autoregressive noise channel models capture the nonlinearities and the first and second order noise statistics of magnetic recording channels. Recently, an iterative Monte Carlo algorithm to compute tight lower bounds on the capacity of intersymbol interference channels was proposed. This algorithm is readily applied to compute capacity lower bounds of data-dependent autoregressive noise channel models. In this work, we extract data-dependent autoregressive noise channel models from simulated micromagnetic perpendicular recording channel waveforms, and then optimize the source distribution to compute capacity lower bounds of the channel models. We note that the lower bound is a bound on the *model*. Finding an accurate model that fits more than just first and second order statistics is still an open problem. The method in this paper is an accurate estimate of the magnetic recording channel capacity only if the channel may be represented by a finite-state machine with up to 2^{10} states (for computational complexity reasons), which applies only to magnetic recording at symbol densities less than 4 symbols per PW50.

1 Introduction

The data-dependent autoregressive (AR) noise channel [1] is an intersymbol interference (ISI) channel, whose signal and correlated noise are both dependent on the channel input pattern. A media noise dominated magnetic recording channel can be accurately modeled as a data-dependent finite-state machine observed through a memoryless channel. As shown in [2, 3], the data-dependent AR noise model is able to readily model the signal nonlinearities and data dependent media noise correlations in magnetic recording channels.

In 2001, Monte Carlo methods [4, 5, 6] were developed to optimize the Markov source distribution and compute achievable information rates for ISI channels. Using this technique, we compute capacity lower bounds of the data-dependent AR noise models extracted from magnetic recording channels. We note that these lower bounds of the models are only estimates of the real capacities because the model captures only the first and second order statistics. The construction of a model that captures more than just first and second order statistics as well as the computation of the capacity of a real channel is still an open problem. By computing

*This work was supported by the National Science Foundation under Grant No. CCR-9904458 and by the National Storage Industry Consortium.

the information density along the recording track, the optimal symbol separation [7] for the recording system is selected to achieve the best trade-off between the system complexity and the linear information density.

Structure: We briefly describe the data-dependent AR noise model in Section 2. In Section 3, we express the mutual information rate in a form suitable for implementing its optimization. In Section 4, we describe the Monte Carlo algorithm used to optimize the Markov channel input distribution. In Section 5, we fit the data-dependent AR noise model to simulated perpendicular magnetic recording channels. Using the Monte Carlo optimization algorithm, we optimize mutual information rates to lower bound the channel capacities of the models.

Notation: The superscript $^{\mathrm{T}}$ denotes matrix and vector transposition. Random variables are denoted by uppercase letters, while their realizations are denoted by lowercase letters. If a random variable is a member of a random sequence, an index $t \in \mathbb{Z}$ is used to denote the discrete time, e.g., X_t. A vector of random variables $[X_i, X_{i+1}, \ldots, X_j]^{\mathrm{T}}$ is shortly denoted by X_i^j, while its realization is shortly denoted by x_i^j. The letter $\mathcal{I}$ denotes the mutual information rate, while the letters H and h denote the entropy and differential entropy, respectively.

2 Data-Dependent AR Signal/Noise Model

The data-dependent AR noise channel model is essentially the same as in [2]. Assuming that the binary channel input sequence is X_k, whose realization is x_k, the channel output sequence is modeled as

$$z_k = y(x_{k+I_2}^{k+I_1}) + n_k. \tag{1}$$

Here, I_1, I_2 are integers and $I_1 \geq I_2$; the value $y(x_{k+I_2}^{k+I_1})$ is the noiseless channel output, which is a function of the $I_1 - I_2 + 1$ input symbols $x_{k+I_2}^{k+I_1}$; the integer $I_1 - I_2 + 1$ is the signal memory window size; and n_k is a sample of an additive Gaussian noise sequence. To capture nonlinearities, we assume that $y(x_{k+I_2}^{k+I_1})$ is constructed as a look-up table rather than the more common convolution of the input symbols and a partial response polynomial.

The noise n_k is modeled 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, i.e.,

$$n_k = \sum_{i=1}^{L} b_i(x_{k+D_2}^{k+D_1}) n_{k-i} + \sigma(x_{k+D_2}^{k+D_1}) w_k, \tag{2}$$

where D_1, D_2 are integers and $D_1 \geq D_2$. The tap-weights $b_i(x_{k+D_2}^{k+D_1})$ and the standard deviation $\sigma(x_{k+D_2}^{k+D_1})$ at time k depend on the $D_1 - D_2 + 1$ input symbols $x_{k+D_2}^{k+D_1}$. Here, $D_1 - D_2 + 1$ is the data-dependent window size of the AR filter, and L is the Markov memory length of the noise. The coefficients $b_i(x_{k+D_2}^{k+D_1})$ and $\sigma(x_{k+D_2}^{k+D_1})$ are stored in look-up tables, and thus are nonlinear functions of $x_{k+D_2}^{k+D_1}$. Being modeled as in (2), the noise sequence is both data-dependent and correlated. Due to the AR filter structure, the noise sample n_k at time k depends on all the previous input symbols up to time $k + D_1$, i.e., $x_{-\infty}^{k+D_1}$. However, if the previous L noise samples n_{k-L}^{k-1} are known, this dependence reduces to the dependence on $x_{k+D_2}^{k+D_1}$ as in (2).

The parameters (I_1, I_2, D_1, D_2, L) specify the *size* of the model. If we define integers M_1 and M_2 to be

$$M_1 = \max(I_1, D_1), \tag{3}$$

$$M_2 = \min(I_2 - L, D_2), \tag{4}$$

then the channel output Z_k at time k depends on the input sequence $X_{k+M_2}^{k+M_1}$, and the total memory length of this model is $M = M_1 - M_2 + 1$. If $M_1 > 0$, the channel output Z_k at time k depends on the future channel inputs $X_{k+1}^{k+M_1}$, which reflects the fact that the bits written on the magnetic recording medium affect the previously written patterns. Assuming that i is the value of the integer defined by the binary input string x_{k+M2+1}^{k+M1}, we let $s_k = i$ be the *state* of the channel at time k. Here, s_k satisfies $0 \leq s_k \leq 2^{M-1} - 1$. The state sequence is a random process which we denote by S_k. If the initial state S_0 of the channel is given, then the channel state sequence S_k and the channel input sequence X_k uniquely determine each other. Thus, the conditional probability density function for the channel output can be written as

$$f\left(z_k | z_{-\infty}^{k-1}, x_{-\infty}^{k+M_1}\right) = f\left(z_k | z_{-\infty}^{k-1}, s_{-\infty}^{k}\right) = f\left(z_k \left| z_{k-L}^{k-1}, s_{k-1}^{k}\right.\right). \tag{5}$$

As shown in [2], given a fixed model size (I_1, I_2, D_1, D_2, L), we can estimate the parameters of the model from a sufficiently long ($K \geq 10^6$) channel input sequence $x_{1+M_2}^{K+M_1}$ and its associated channel output sequence z_1^K. For each possible $(I_1 - I_2 + 1)$-long channel input sequence pattern $x_{k+I_2}^{k+I_1} = \alpha$, we estimate the noiseless channel output look-up table function $y(\alpha)$ as

$$\hat{y}(\alpha) = \frac{1}{K_\alpha} \sum_{k: x_{k+I_2}^{k+I_1} = \alpha} z_k. \tag{6}$$

Here, K_α is the number of the occurrences of the pattern $x_{k+I_2}^{k+I_1} = \alpha$ in the channel input sequence $x_{1+M_2}^{K+M_1}$.

Similarly, for each possible $(D_1 - D_2 + 1)$-long channel input sequence pattern $x_{k+D_2}^{k+D_1} = \beta$, we estimate the data-dependent noise covariance $c_{i,j}(\beta) = \mathrm{E}\left[N_{k-i} N_{k-j} \left| x_{k+D_2}^{k+D_1} = \beta\right.\right]$ as

$$\hat{c}_{i,j}(\beta) = \frac{1}{K_\beta} \sum_{k: x_{k+D_2}^{k+D_1} = \beta} \left(z_{k-i} - \hat{y}\left(x_{k-i+I_2}^{k-i+I_1}\right)\right)\left(z_{k-j} - \hat{y}\left(x_{k-j+I_2}^{k-j+I_1}\right)\right). \tag{7}$$

Here, K_β is the number of occurrences of the pattern $x_{k+D_2}^{k+D_1} = \beta$ in the channel input sequence $x_{1+M_2}^{K+M_1}$. Then we can find the estimates $\hat{b}_i(\beta)$ and $\hat{\sigma}(\beta)$ of the data-dependent AR filter coefficients $b_i(\beta)$ and the noise variance $\sigma(\beta)$ by solving the following augmented Yule-Walker equations (Wiener-Hopf equations)

$$\hat{\mathbf{C}}(\beta) \begin{bmatrix} \hat{b}_1(\beta) \\ \hat{b}_2(\beta) \\ \vdots \\ \hat{b}_L(\beta) \end{bmatrix} = \begin{bmatrix} -\hat{\sigma}^2(\beta) + \hat{c}_{0,0}(\beta) \\ \hat{c}_{0,1}(\beta) \\ \vdots \\ \hat{c}_{0,L}(\beta) \end{bmatrix}, \tag{8}$$

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

$$\hat{\mathbf{C}}(\beta) = \begin{bmatrix} \hat{c}_{1,0}(\beta) & \hat{c}_{2,0}(\beta) & \dots & \hat{c}_{L,0}(\beta) \\ \hat{c}_{1,1}(\beta) & \hat{c}_{2,1}(\beta) & \dots & \hat{c}_{L,1}(\beta) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}_{1,L}(\beta) & \hat{c}_{2,L}(\beta) & \dots & \hat{c}_{L,L}(\beta) \end{bmatrix}. \tag{9}$$

The look-up functions $\hat{y}(\alpha)$, $\hat{b}_i(\beta)$, and $\hat{\sigma}(\beta)$ capture signal nonlinearities and the data-dependent noise correlation in the real channel, and match the channel output waveform to the first and second order statistics. Larger model sizes usually match the real channel with more accuracy. For an appropriate model size (I_1, I_2, D_1, D_2, L), the mismatch of this model is proportional to the "level" of noise non-Gaussianity in the real channel [2].

3 Mutual Information Rate

The channel model in Section 2 is an indecomposable finite-state machine channel. Given the initial state of the channel S_0, the state sequence S_k and the channel input sequence X_k determine each other uniquely. The mutual information rate is not affected by the initial state of the channel, and may be expressed as

$$\mathcal{I}(X;Z) = \lim_{K\to\infty} \frac{1}{K} I\left(X_1^K; Z_1^K | S_0, Z_{1-L}^0\right) = \lim_{K\to\infty} \frac{1}{K} I\left(S_1^K; Z_1^K | S_0, Z_{1-L}^0\right), \tag{10}$$

where $I(A;B|C)$ denotes the mutual information between random variables A and B, given C.

We assume that the source S_k is a stationary Markov process with memory length $M = M_1 - M_2 + 1$. Note that M is also the memory length of the channel model in Section 2. At any time k, this Markov source is defined by the transition probabilities

$$P_{ij} = \Pr\left(S_k = j | S_{k-1} = i\right). \tag{11}$$

A transition from state i to state j is considered to be *invalid* if the Markov state sequence cannot be taken from state i to state j. The transition probability P_{ij} for an invalid transition is thus zero. A *valid* transition is a transition that is not invalid. A trellis section, denoted by $\mathcal{T}$, is the set of all valid transitions. Thus, a valid transition (i,j) satisfies $(i,j) \in \mathcal{T}$.

We use μ_i to represent the Markov steady state probability, i.e.,

$$\mu_j = \lim_{t\to\infty} \Pr\left(S_t = j\right) = \sum_i \mu_i P_{ij}. \tag{12}$$

For this stationary Markov source, using the chain rule, we can rewrite the mutual information rate in a form suitable for its maximization

$$\begin{aligned}
\mathcal{I}(X;Z) &= \lim_{K\to\infty} \frac{1}{K} I\left(S_1^K; Z_1^K | S_0, Z_{1-L}^0\right) && (13)\\
&= \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} I\left(S_k; Z_1^K | S_0^{k-1}, Z_{1-L}^0\right) && (14)\\
&= \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} I\left(S_k; Z_1^K | S_{k-1}, Z_{1-L}^0\right) && (15)\\
&= \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \left[H\left(S_k | S_{k-1}, Z_{1-L}^0\right) - H\left(S_k | S_{k-1}, Z_{1-L}^K\right)\right] && (16)\\
&= \sum_{i,j} \mu_i P_{ij} \left(\log \frac{1}{P_{ij}} + T_{ij}\right). && (17)
\end{aligned}$$

Similar to [6], the term T_{ij} in (17) is defined as

$$T_{ij} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \left\{ \mathrm{E}_{Z_{1-L}^K | i,j} \left[\log P_k\left(i,j|Z_{1-L}^K\right)\right] - \mathrm{E}_{Z_{1-L}^K | i} \left[\log P_{k-1}\left(i|Z_{1-L}^K\right)\right] \right\}. \tag{18}$$

Here, $P_k\left(i|Z_{1-L}^K\right)$ and $P_k\left(i,j|Z_{1-L}^K\right)$ stand for the conditional probabilities $\Pr\left(S_k = i|Z_{1-L}^K\right)$ and $\Pr\left(S_{k-1} = i, S_k = j|Z_{1-L}^K\right)$, respectively. $\mathrm{E}_{Z_1^K|i,j}$ denotes the conditional expectation

taken over the variable Z_1^K when the pair (S_{k-1}, S_k) equals (i, j). Similarly, $\mathrm{E}_{Z_1^K|i}$ is the conditional expectation taken over the variable Z_1^K when $S_{k-1} = i$.

As shown in [6], we can estimate T_{ij} as the empirical expectation

$$\hat{T}_{ij} = \frac{1}{K} \sum_{k=1}^{K} \left[\log \frac{P_k\left(i, j | z_{1-L}^K\right)^{\frac{P_k\left(i,j|z_{1-L}^K\right)}{\mu_i P_{ij}}}}{P_{k-1}\left(i | z_{1-L}^K\right)^{\frac{P_{k-1}\left(i|z_{1-L}^K\right)}{\mu_i}}} \right]. \tag{19}$$

where the probabilities $P_k\left(i|Z_{1-L}^K\right)$ and $P_k\left(i, j|Z_{1-L}^K\right)$ are computed by running the sum-product (BCJR) algorithm [8]. Assuming the channel output process is ergodic, $\hat{T}_{ij}$ converges to T_{ij} with probability 1.

4 Information Rate Maximization Algorithm

In [6], an iterative Monte Carlo algorithm was proposed to compute the optimal stationary Markov source transition probabilities (11) that maximize the mutual information rate (17). The same algorithm applies to data-dependent AR noise channels.

Algorithm 1 A STOCHASTIC METHOD FOR OPTIMIZING MARKOV PROCESS TRANSITION PROBABILITIES

Initialization: Pick an arbitrary distribution P_{ij} that satisfies the following two constraints:
1) if$(i, j) \in \mathcal{T}$ then $0 < P_{ij} < 1$.
2) for any i, require that $\sum_j P_{ij} = 1$.

Repeat until convergence

Step 1: For K large, generate s_1^K according to the transition probabilities P_{ij} and pass them through the noisy data-dependent AR noise channel model in Section 2 to get z_1^K.

Step 2: Run the forward-backward sum-product (BCJR) algorithm [8] and compute $\hat{T}_{ij}$ according to (19).

Step 3: Estimate the noisy adjacency matrix as

$$\hat{A}_{ij} = \begin{cases} 2^{\hat{T}_{ij}}, & \text{if}(i, j) \in \mathcal{T} \\ 0, & \text{otherwise} \end{cases}, \tag{20}$$

and find its maximal real eigenvalue $\hat{\lambda}_{max}$ and the corresponding eigenvector

$$\underline{\hat{v}} = [\hat{v}_0, \hat{v}_1, \ldots, \hat{v}_{V-1}]^{\mathrm{T}}, \tag{21}$$

where $V = 2^{M-1}$.

Step 4: Compute the entries of the new transition probability matrix for $(i, j) \in \mathcal{T}$ as

$$P_{ij} = \frac{\hat{v}_j \hat{A}_{ij}}{\hat{v}_i \hat{\lambda}_{max}}. \tag{22}$$

end

Using the Markov transition probabilities obtained by Algorithm 1, we can compute the mutual information rate (17) as a channel capacity lower bound.

5 Applications to recording channels

First, we fit data-dependent AR noise channel models of Section 2 to simulated perpendicular magnetic recording channels. Then we obtain channel capacity lower bounds of the models using Algorithm 1 from Section 4. This way, we estimate the capacity of the perpendicular magnetic recording channel.

5.1 Perpendicular microtrack model

We use a microtrack model as in Figure 1 to simulate the perpendicular recording channel. This model is similar to [9], with the only difference being that we assume a perpendicular channel impulse response rather than a longitudinal transition response.

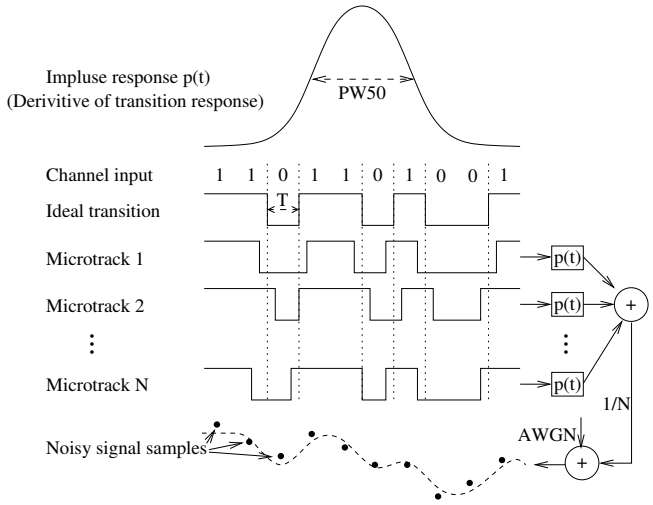

Figure 1: A microtrack perpendicular recording channel model

The recording track is modeled to consist of N parallel microtracks. The number of the microtracks N models the width of the recording track. Each microtrack is one subchannel. The input signals to these microtrack subchannels are the rectangular waveforms whose two amplitudes correspond to the input symbols 1 and 0, respectively. Let T be the symbol separation, which is the ideal (or average) duration of one binary symbol in the waveform. For each microtrack subchannel, the media noise is modeled by the transition jitter noise, i.e., transitions between 1's and 0's drift around their ideal positions by an i.i.d. Gaussian random distance whose mean is zero and variance is σ_j^2. Two adjacent transitions in a microtrack cancel each other if the distance between them is less than a threshold T_{min}, as modeled in [9]. The random jitters are assumed to be independent for different microtrack subchannels. The channel impulse response for each microtrack is approximated by a Gaussian function $p(t)$ parameterized by the half-amplitude width $PW50$. The parameter $PW50$ is affected by the quality of the reading head. The ratio between $PW50$ and T corresponds to the level of intersymbol interference (ISI). The channel output of the perpendicular recording channel is the sum of the subchannel outputs and the additive white Gaussian noise (AWGN) with variance σ_n^2 (or power spectral density N_0).

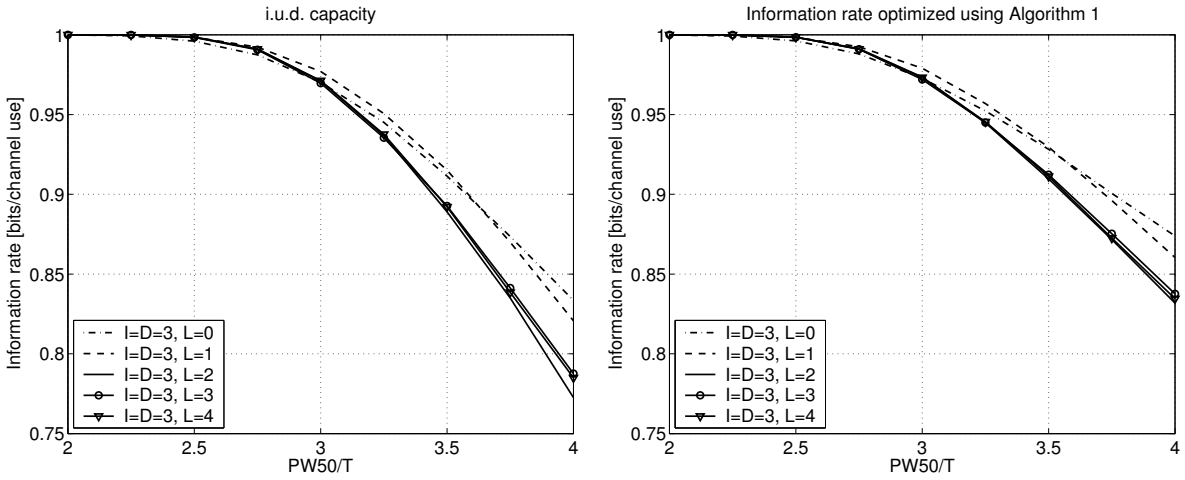

Figure 2: Capacity estimates for varying values of the noise memory L in the AR model ($I_1 = -I_2 = I = D = D_1 = -D_2$)

5.2 Channel capacity estimation and the AR model size

In our simulations, we fix the parameters $N = 10$, $PW50 = 140[\mathrm{nm}]$, $T_{min} = 31.5[\mathrm{nm}]$, and $\sigma_j^2 = 537.6[\mathrm{nm}^2]$. We do not claim that these are parameters used in any product. They were simply chosen as a working point to illustrate the effects on information rates. The AWGN variance σ_n^2 keeps a linear relationship with the sampling rate $1/T$, that is $\sigma_n^2 \propto 1/T$. We set N_0 so that the AWGN variance σ_n^2 is 10% of the total noise power when $T = 46.7[\mathrm{nm}]$. Here media noise is defined to be the difference between the channel output when $N = 10$ tracks are read without AWGN and the channel output when $N = \infty$ tracks are read without AWGN. For this recording system setup, we vary T so that the ratio $PW50/T$ changes from 2.0 to 4.0.

We fit data-dependent AR noise models with different model sizes (I_1, I_2, D_1, D_2, L) to the simulated perpendicular recording channel waveforms. Then, for each of the extracted data-dependent AR noise models, we compute the i.u.d. channel capacity, which is defined as the information rate when the inputs are independent and uniformly distributed (i.u.d.) binary random variables. For each of the extracted data-dependent AR noise models, we also compute a capacity lower bound, which is the achievable information rate when the source is optimized by Algorithm 1 in Section 4. We note that the i.u.d. capacity is the rate limit for random linear error correcting codes, e.g., turbo codes [10] with random interleavers and random LDPC codes [11], while the channel capacity is achieved by specially shaped (possibly nonlinear) codes [12].

Usually, a data-dependent AR noise model of a bigger size describes the real channel more accurately, and its capacity is closer to the real capacity. We note that models with big enough sizes should have similar capacities. Figure 2 shows the channel capacity estimates of a series of models which have the same data-dependent window size but different noise autoregressive degrees, i.e., $I_1 = -I_2 = D_1 = -D_2 = 3$ and $L = 0, 1, \ldots, 4$. By comparing the curves, we can see that the capacity estimates converge to a single curve when the noise autoregressive degree L is increased. The required value of L to achieve the convergence increases with the ratio $PW50/T$. That is, at $PW50/T = 2$ a noise memory length $L = 0$ suffices to achieve convergence of the capacity estimates, while at $PW50/T = 4$, the noise memory must be increased to $L = 3$. For the ratios $PW50/T$ less than 4.0, Figure 2 show that $L = 3$ is enough to achieve convergence of the curves to a single curve.

In Figure 3, the data-dependent window size $I_1 - I_2 + 1 (= D_1 - D_2 + 1)$ varies from

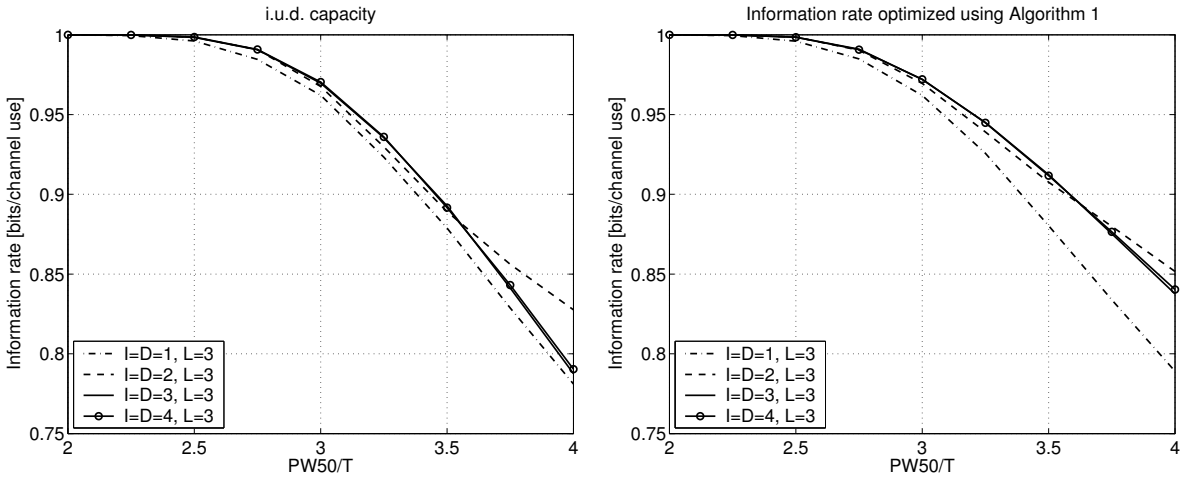

Figure 3: Capacity estimates for varying values of the data-dependent window size $I_1 - I_2 + 1 (= D_1 - D_2 + 1)$ in the AR model ($I_1 = -I_2 = I = D = D_1 = -D_2$)

3 symbols to 9 symbols while the noise memory length is fixed to $L = 3$. The curves show that the required data-dependent window size to achieve convergence of the capacity estimates is proportional to the ratio $PW50/T$. For the ratios $PW50/T$ less than 4.0, we can see that $I_1 = -I_2 = D_1 = -D_2 = 3$ is enough to achieve the convergence of the curves. Combining Figure 2 and Figure 3, we conclude that the model size $I_1 = -I_2 = D_1 = -D_2 = L = 3$ is enough to estimate the capacity bounds of the simulated recording channels for $PW50/T \leq 4.0$.

5.3 The optimal symbol separation

An important issue for designing the recording system is to select the optimal symbol separation T, so that the information stored in every unit length of recording track is maximized. In [7], this optimization was done for a simulated longitudinal recording channel with white noise. Ryan [7] observed that the linear information density (computed as the i.u.d. capacity divided by the symbol separation) is maximized by an optimal symbol separation. The optimal symbol separation exists for the following two reasons. 1) For a very large symbol separation T, the i.u.d. capacity is usually very close to $1[\mathrm{bit/channel\ use}]$, and thus its linear density decreases approximately linearly with $1/T$. 2) For a very small symbol separation T, the intersymbol interference (ISI) significantly degenerates the i.u.d. capacity while the white noise increases since it is proportional to $1/T$, so the linear information density also becomes very small. However, there doesn't seem to exist an optimal symbol separation to maximize the linear information density when it is computed as the *optimized* information rate in $[\mathrm{bits/channel\ use}]$ divided by the symbol separation T. A smaller symbol separation always delivers a higher optimized linear information density. This means that in order to maximize the linear information density, theoretically, we should always require a smaller symbol separation T. However, reducing T increases the ISI memory and thus exponentially increases the required channel model size (trellis complexity). Systems with extremely high trellis complexity are not implementable. For this reason, in order to find the optimal symbol separation T, we must consider what is practically feasible, and find a good trade-off between the linear information density and feasibility.

Figure 4 depicts the linear information density (in $[\mathrm{kbits/inch}]$) for both the i.u.d. source and the optimized source and also shows the code rates (r) in $[\mathrm{bits/channel\ use}]$ needed to

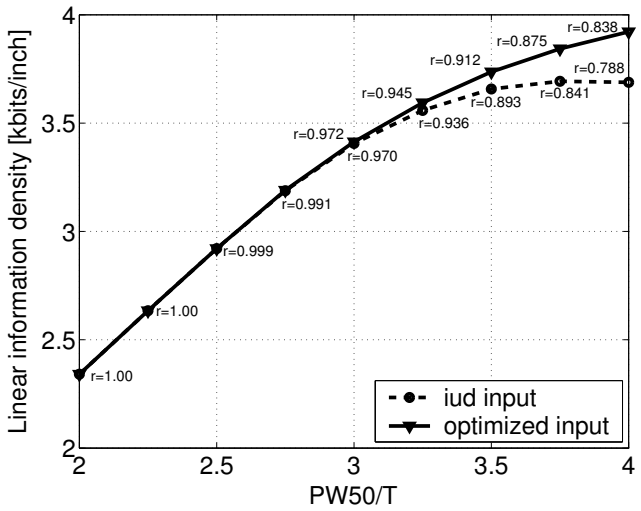

Figure 4: Linear information density estimates

achieve these linear information densities. As seen in Figure 4, the maximum of the linear information density for the i.u.d. source is achieved at $PW50/T \approx 3.8$, which leads to a symbol separation $T \approx 37[\mathrm{nm}]$ and an linear information density of $3.7[\mathrm{kbits/inch}]$. The corresponding code rate is $r = 0.84[\mathrm{bits/channel\ use}]$. If we consider a realistic system complexity requirement: the size of the finite state machine channel model should be less than 2^{10} and the code rate needs to be higher than $0.9[\mathrm{bits/channel\ use}]$), then the linear information density for the i.u.d. source is maximized at $PW50/T \approx 3.4$, which leads to a symbol separation $T \approx 41[\mathrm{nm}]$ and a linear information density of $3.6[\mathrm{kbits/inch}]$. Under the same constraint, the maximum of the linear information density for the optimized source is achieved at $PW50/T \approx 3.6$, which leads to an optimal symbol separation $T \approx 39[\mathrm{nm}]$ and a linear information density of $3.8[\mathrm{kbits/inch}]$. However, in order to achieve this linear information density computed for the optimized source, we cannot use a random linear code, but must use a code that shapes the channel input to fit the channel [12].

6 Conclusion

We fitted data-dependent AR noise models to the microtrack perpendicular recording channel, and then computed the capacity bounds of the models. This way, we estimated the capacity of the real channel. By estimating the capacity of the recording channel, we selected the optimal symbol separation to meet the best trade-off between the system feasibility and the linear information density stored on the recording track.

References

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

[2] A. Kavčić and M. Srinivasan, "The minimum description length principle for modeling recording channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 719–729, April 2001.

[3] J. Stander and A. Patapoutian, "Performance of a Signal-dependent Autoregressive channel model," in *IEEE Transactions on Magnetics*, vol. 36, pp. 2197-2199, Sept. 2000.

[4] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proceedings IEEE International Conference on Communications 2001*, (Helsinki, Finland), June 2001.

[5] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proceedings IEEE Global communications Conference 2001*, (San Antonio, Texas), November 2001.

[6] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *Proceedings IEEE Global Communications Conference 2001*, (San Antonio, Texas), November 2001. available at http://hrl.harvard.edu/~kavcic/publications.html.

[7] W. Ryan, "Optimal code rates for Lorentzian channel models," *submitted to the 2003 IEEE Int. Conf. Comm.*.

[8] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Sept. 1974.

[9] 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. MAG-27, pp. 4978–4980, Nov. 1991.

[10] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. on Communications*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[11] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1962.

[12] X. Ma, N. Varnica, and A. Kavčić, "Matched information rate codes for binary ISI channels," in *Proc. IEEE Int. Symp. on Information Theory*, (Laussane, Switzerland), July 2002.