On the Achievable Information Rates of Finite State ISI Channels

Henry D. Pfister, Joseph B. Soriaga, and Paul H. Siegel
Department of Electrical and Computer Engineering, 0407
University of California, San Diego, La Jolla, CA 92093-0407

***Abstract*— In this paper, we present two simple Monte Carlo methods for estimating the achievable information rates of general finite state channels. Both methods require only the ability to simulate the channel with an a posteriori probability (APP) detector matched to the channel. The first method estimates the mutual information rate between the input random process and the output random process, provided that both processes are stationary and ergodic. When the inputs are i.i.d. equiprobable, this rate is known as the Symmetric Information Rate (SIR). The second method estimates the achievable information rate of an explicit coding system which interleaves m independent codes onto the channel and employs multistage decoding. For practical values of m, numerical results show that this system nearly achieves the SIR. Both methods are applied to the class of partial response channels commonly used in magnetic recording.**

I. Introduction

Determining the achievable transmission rates of information across noisy channels has been one of the central pursuits in information theory since Shannon invented the subject in 1948. In particular, Gallager derived a general formula for the capacity of a finite state (FS) channel in [1, p.100]. This formula is not always easily evaluated, especially in cases where the receiver uses imperfect channel state estimates derived from previously received symbols. Goldsmith and Varaiya give some interesting results for the case where these estimates are independent of the transmitted sequence in [2]. We focus on the FS intersymbol interference (FSISI) channel, where the imperfect channel state estimates are dependent on the transmitted sequence. Recently, this topic has become very popular and results similar to ours were obtained independently and reported in [3] by Arnold and Loeliger and in [4] by Sharma and Singh.

A very common subclass of the FSISI channel is the linear filter channel with additive white Gaussian noise (AWGN), which is described by

$$y_k = \sum_{i=0}^{\nu} h_i x_{k-i} + n_k, \tag{1}$$

where $\{x_k\}$ is the channel input (taken from a discrete alphabet), $\{y_k\}$ is the channel output, and $\{n_k\}$ is i.i.d., zero mean Gaussian noise. For this channel, the Symmetric Information Rate (SIR), which is the maximum information rate achievable with i.i.d. equiprobable inputs, has been considered by many authors. Shamai *et al.* derived a number of analytical bounds in [5], and Hirt developed a Monte Carlo method for estimating the SIR in [6]. Computing the capacity of this channel is more difficult, however, because it requires maximizing the achievable information rate over all input distributions.

This work was supported in part by the National Science Foundation under grant NCR-9612802, the National Storage Industry Consortium, by UC MICRO Grant 99-110 in cooperation with Marvell Semiconductor, Inc., and by the University of California Office of the President.

In this paper, we develop two methods of estimating the achievable information rates of FSISI channels. The complexity of each method is equivalent to the complexity of simulating an *a posteriori* probability (APP) detector matched to the channel for very long input sequences. The first method estimates the mutual information rate between the input random process and the output random process, provided that both processes are stationary and ergodic. It is general enough to allow the mutual information to be maximized over Markov input distributions of increasing length, and thus can be used to estimate a sequence of non-decreasing lower bounds on capacity. Although it was derived independently, it is identical to a method proposed in both [3] and [4]. The second method estimates the achievable information rate of a particular coding system, where the encoder interleaves m independent codes onto the channel and the decoder employs multistage decoding. On a more practical note, numerical results show the second method nearly achieves the SIR for moderate values of m.

II. Achievable Rates

A. General Ergodic Channels

We can define any communication channel using the conditional density function of the channel output vector given the channel input vector. In general, the input vector, $\mathbf{x} = (x_1, \ldots, x_n)$ with each $x_i \in \mathbb{X}$, is a realization of the random process $\mathbf{X} = (X_1, \ldots, X_n)$. Likewise, the output vector, $\mathbf{y} = (y_1, \ldots, y_n)$ with each $y_i \in \mathbb{Y}$, is the realization of a random process $\mathbf{Y} = (Y_1, \ldots, Y_n)$. We denote subvectors using the notation $\mathbf{x}_i^j = (x_i, \ldots, x_j)$ and $\mathbf{X}_i^j = (X_i, \ldots, X_j)$. The channel is completely defined by $\mathbb{X}$, $\mathbb{Y}$, and the conditional density function $f_{\mathbf{y}}(\mathbf{y}|\mathbf{x})$. The mutual information rate between the input random process, $\mathbf{X}$, and the output random process, $\mathbf{Y}$, is

$$I(\mathcal{X};\mathcal{Y}) = \lim_{n\to\infty} \frac{1}{n} I(\mathbf{X}_1^n; \mathbf{Y}_1^n), \tag{2}$$

and it is the maximum achievable information rate for this channel and input distribution. Unfortunately, this quantity is not well defined for all channels and input processes, so we limit our discussion to stationary ergodic channels with stationary ergodic inputs. Entropy rates are well defined for stationary ergodic random processes (SERPs), and we define a stationary ergodic channel as any channel whose output is a SERP whenever its input is a SERP.

The sample entropy rate of a SERP is the random variable, $\hat{H}_n(\mathcal{Y}) = -(1/n)\log Pr(\mathbf{Y}_1^n)$, which converges almost surely to the true entropy rate. Mathematically speaking, we have

$$\lim_{n\to\infty} -\frac{1}{n}\log Pr(\mathbf{Y}_1^n) = H(\mathcal{Y}), \quad \text{a.s.}$$

which is known as the Shannon-McMillan-Breiman theorem for discrete SERPs [7, p. 474]. For continuous SERPs, this result was extended to the density function $f(\mathbf{Y}_1^n)$ and the differential entropy rate $h(\mathcal{Y})$ by Barron [8]. In either case, these results can be used to estimate the entropy of any SERP, and thus estimate the mutual information rate, $I(\mathcal{X};\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X})$, of any stationary ergodic channel. Consequently, the complexity of estimating the mutual information rate is essentially equal to the complexity of generating a long realization, $\mathbf{y}_1^n$, and computing $\log Pr(\mathbf{y}_1^n)$. This complexity can usually be minimized by using the relationship $\log Pr(\mathbf{y}_1^n) = \sum_{i=1}^n \log Pr(y_i|\mathbf{y}_1^{i-1})$.

Finally, all values of (2) calculated for stationary ergodic channels are achievable rates since these theorems extend the Asymptotic Equipartition Property (AEP) to SERPs. Once this is established, the ideas of joint typicality and typical set decoding, as described in [7, p. 194], can be used to prove achievability of the rate $I(\mathcal{X};\mathcal{Y})$.

B. Finite State Channels

A FS channel is defined by its discrete input alphabet $\mathbb{X}$, its discrete (or continuous) output alphabet $\mathbb{Y}$, its finite set of states $\mathbb{Q}$, and its transition probability (or density) function $Pr(Y_i, Q_{i+1}|X_i, Q_i)$. If we assume that the input vector $\mathbf{X}_1^n$ is also generated by a finite state machine, then we can view the channel output as coming from a finite state Hidden Markov Model (HMM) with observed outputs $\mathbf{Y}_1^n$ and hidden state sequence $\mathbf{Q}_1^{n+1}$. Using the finite state property of the input process, we can choose the state space to be large enough so that the input distribution satisfies $Pr(X_i|\mathbf{X}_1^{i-1}, Q_1) = Pr(X_i|Q_i)$. We note that the distributions $Pr(Y_i, Q_{i+1}|X_i, Q_i)$ and $Pr(X_i|Q_i)$ are assumed to be stationary and are simply indexed by i for clarity.

Consider any channel with i.i.d. noise whose ISI is due to a finite impulse response (FIR) linear filter (e.g., the channel defined by (1)). In this case, the channel state is defined by the previous ν inputs. We also assume that the state of the input process is defined by its previous κ outputs; this allows us to choose $\eta = \max(\nu, \kappa)$, and let the state variable $Q_i = \{\mathbf{X}_{i-\eta}^{i-1}\}$ without loss of generality.

We restrict our attention to stationary ergodic FSISI channels and apply the Shannon-McMillan-Breiman theorem (or Barron's extension) to these channels in an efficient manner. We can generate realizations, $\mathbf{y}_1^n$, of the output random process and apply an APP detector matched to the channel to compute the sample entropy rate, $\hat{H}_n(\mathcal{Y}) = -(1/n)\log Pr(\mathbf{y}_1^n)$.

Recall that the forward pass of the APP algorithm computes $Pr(Q_i|\mathbf{y}_1^{i-1})$ recursively. Using the definition, $\alpha_i(q) = Pr(Q_i = q|\mathbf{y}_1^{i-1})$, we have the forward recursion

$$\alpha_{i+1}(q) = \sum_{q'\in\mathbb{Q}} \alpha_i(q') Pr(Q_{i+1} = q|Q_i = q', y_i).$$

TABLE I

THE TRANSFER FUNCTION (TF) AND NORMALIZED RESPONSE (NR) OF VARIOUS PARTIAL RESPONSE CHANNELS

Channel	TF: $G(D)$	NR
Dicode	$(1-D)$	[1 -1]/$\sqrt{2}$
EPR4	$(1-D)(1+D)^2$	[1 1 -1 -1]/2
E^2PR4	$(1-D)(1+D)^3$	[1 2 0 -2 -1]/$\sqrt{10}$

This uses the identity, $Pr(Q_{i+1}|Q_i, \mathbf{Y}_1^i) = Pr(Q_{i+1}|Q_i, Y_i)$, which is implied by the conditional independence of $\mathbf{Y}_1^{i-1}$ and Q_{i+1} given Q_i. The sample entropy, $\hat{H}_i(\mathcal{Y})$, is defined by the recursion $\hat{H}_{i+1}(\mathcal{Y}) = \frac{i}{i+1}\hat{H}_i(\mathcal{Y}) + \frac{1}{i+1}\Delta\hat{H}_i(\mathcal{Y})$, where

$$\Delta\hat{H}_i(\mathcal{Y}) = -\log \sum_{q\in\mathbb{Q}} \alpha_{i+1}(q) Pr(y_{i+1}|Q_{i+1} = q).$$

This is a natural extension of the recursion for $Pr(\mathbf{y}_1^n)$,

$$Pr(\mathbf{y}_1^{i+1}) = \sum_{q\in\mathbb{Q}} Pr(\mathbf{y}_1^i)\alpha_{i+1}(q) Pr(y_{i+1}|Q_{i+1} = q).$$

A similar recursion can be derived for $\hat{H}_n(\mathcal{Y}|\mathcal{X}) = -(1/n)\log Pr(\mathbf{y}_1^n|\mathbf{x}_1^n)$. For channels whose state transitions are determined only by the current state and input, we instead evaluate the limiting value, $H(\mathcal{Y}|\mathcal{X}) = \lim_{n\to\infty}\frac{1}{n}H(\mathbf{Y}_1^n|\mathbf{X}_1^n)$. This can be written analytically as

$$H(\mathcal{Y}|\mathcal{X}) = -\sum_{q,q'\in\mathbb{Q}} \pi(q)p(q'|q)\int_{\mathbb{Y}} p(y|q',q)\log p(y|q',q)dy,$$

where $\pi(q) = \lim_{i\to\infty} Pr(Q_i = q)$, $p(y|q',q) = Pr(Y_i = y|Q_{i+1} = q', Q_i = q)$, and $p(q'|q) = Pr(Q_{i+1} = q'|Q_i = q)$.

The final step in this process is computing the estimate of the mutual information rate,

$$\hat{I}_n(\mathcal{X};\mathcal{Y}) = \hat{H}_n(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}). \tag{3}$$

For many channels, we can show that $\hat{I}_n(\mathcal{X};\mathcal{Y})$ is asymptotically Gaussian with asymptotic mean $I(\mathcal{X};\mathcal{Y})$ and variance $O(n^{-1/2})$ by first showing that $\hat{I}_n(\mathcal{X};\mathcal{Y})$ is a function of an ergodic Markov chain (with state sequence $\{Q_i, X_i, Y_i, \alpha_i(q)\}$), and then applying an appropriate central limit theorem [9]. We note that simulations are run using an arbitrary initial channel state, Q_1, and then initializing the state probability vector of the APP detector with $\alpha_1(Q_1) = 1$. This gives the APP detector perfect knowledge of the initial state. For large enough n, however, the effect of the initial conditions should be negligible.

Computing the channel capacity involves maximizing the mutual information rate $I(\mathcal{X};\mathcal{Y})$ over all possible input distributions. Let S_η be the set of all Markov input distributions with memory η, which implies that $Pr(X_i|\mathbf{X}_1^{i-1}) = Pr(X_i|\mathbf{X}_{i-\eta}^{i-1})$. The sequence $\{C_\eta\}$, which is defined by

$$C_\eta = \lim_{n\to\infty}\frac{1}{n}\max_{Pr(\mathbf{X})\in S_\eta} I(\mathbf{X}_1^n;\mathbf{Y}_1^n),$$

is a sequence of non-decreasing lower bounds on the true capacity of the channel and can be estimated with (3).

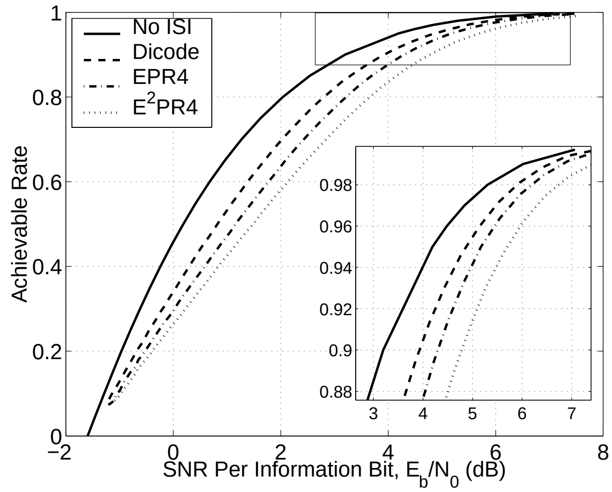

Fig. 1. The SIR for various channels, estimated with $n = 10^7$.

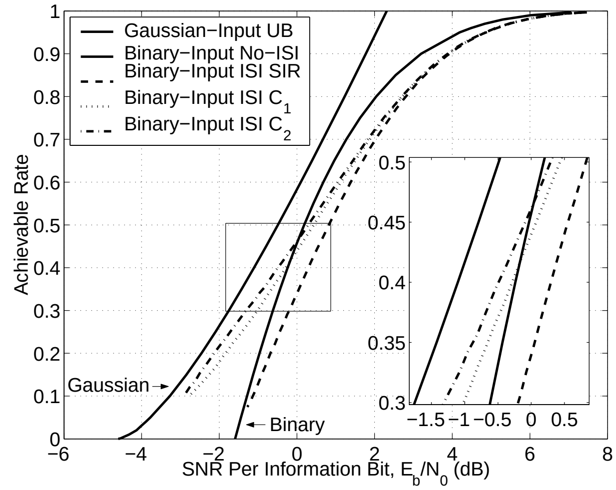

Fig. 2. Achievable rates on the dicode channel with optimized $Pr(\mathbf{X})$.

C. Simulation Results

In these simulations, we restrict our attention to power normalized binary-input channels listed in Table I which are sometimes used to model magnetic recording channels. The results of estimating the SIR for each channel, using the method of Section II-B, are shown in Fig. 1. Notice that the SIR of each ISI channel is worse than that of the binary-input AWGN channel.

The results of using standard gradient based methods to optimize Markov input distributions for the dicode channel are shown in Fig. 2. Our estimates of C_1 and C_2 show a clear improvement over the SIR, and even surpass the capacity of the binary-input AWGN channel at normalized SNRs less than -0.5 dB. We discuss this behavior in the next section.

D. The Low Rate Shannon Limit

The Shannon limit is the minimum SNR per bit required for reliable communication over an AWGN channel at a particular code rate. At low rates, this quantity approaches a limiting value which we call the low rate Shannon limit. It is well-known that the low rate Shannon limit is achieved, for Gaussian-input ISI channels, by concentrating the input signal power sharply around a maximum of the channel's frequency response. We constructively upper bound this limit, for discrete-input ISI channels, by converting the problem into a coding problem for a memoryless binary-input channel. More precisely, we encode the information with a binary code and map each code bit to a p-symbol sequence before transmission. We use a single sequence and its complement for the mapping and tail off each sequence to prevent ISI. Demodulating the p-symbol sequences with a matched filter results in a binary-input (BI) AWGN channel with an improved SNR.

Clearly, one should choose the p-symbol sequence which maximizes the ratio of output power to input power. Given the channel frequency response, $G(f)$, the theoretical maximum of this ratio is $T_{max} = \max_f |G(f)|^2$. For the binary-input dicode channel, the periodic sequence $(1, -1, 1, -1, \dots)$ gives a power ratio of 2 as p goes to infinity. Since $T_{max} = 2$ for this channel, this implies that the low rate Shannon limit is -4.59 dB. Simulation results in Fig. 2 seem to agree with this result, as one can extrapolate the low rate Shannon limit to be somewhere around -4 dB.

III. Coding to Achieve the SIR

A. Interleaving Binary Codes

We now consider a realizable[1] method for achieving the SIR by interleaving m independent random codes, and using an m-stage decoder. At the ith stage, we first decode the channel, using symbol estimates from all previous $(i-1)$ stages as *a priori* information, and then decode the ith code using a maximum likelihood decoder (MLD) designed for memoryless channels. The MLD feeds its estimates of the information and channel bits to subsequent decoding stages. We determine the achievable information rates for each stage by examining the output statistics of the channel APP detector at each stage. The *a priori* information from previous stages allows subsequent stages to achieve higher rates, and the overall code rate is given by

$$R_{av} = \frac{1}{m} \sum_{i=1}^{m} R_i,$$

where $R_1 \leq R_2 \leq \cdots \leq R_m$. We note that this method for determining rates is significantly different from method in Section II-B, but the results are almost identical for large enough m.

For clarity, we focus our analysis on the case of two interleaved codes (i.e., $m = 2$), and continue with a closer examination of the system construction. At the encoder, illustrated in Fig. 3, a block of information bits $\mathbf{u}$ is broken into two smaller information blocks, $\mathbf{a}$ and $\mathbf{b}$, each containing nR_A and nR_B bits, respectively. These information blocks are then encoded into a pair of n-bit codewords, $\mathbf{A}$ and $\mathbf{B}$, which are finally interleaved into a single block of coded bits $\mathbf{X} = (A_1, B_1, A_2, B_2, \dots, A_n, B_n)$ for transmission through the channel. Clearly, this can be viewed as a single encoder for a $(2n\,, nR_A + nR_B)$ binary block code.

[1]This method is realizable in the sense that good coding systems developed for memoryless channels in [10] can be used to approach the SIR.

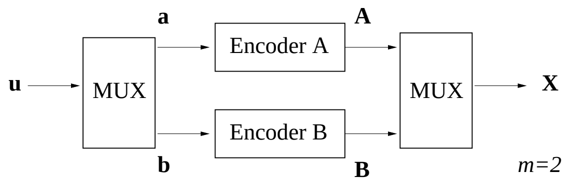

Fig. 3. Encoder for interleaved codes.

The multistage decoder used to recover $\mathbf{u}$ is depicted in Fig. 4. First, (Step 1) an APP detector matched to the channel is applied to the received block $\mathbf{Y}$ to obtain soft estimates, $\mathbf{L}_A$, of the channel inputs $\mathbf{A}$. Since bit values are marginally equiprobable at any position in a random binary code, the first APP detector assumes that channel input bit values are equiprobable at every time index. This is equivalent to the detector receiving no *a priori* information. The Code A MLD (Step 2) then recovers $\hat{\mathbf{a}}$ from the soft estimates $\mathbf{L}_A$, operating as it would on a memoryless channel. In other words, it assumes that $Pr(L_{A,i}, L_{A,j}|A_i, A_j) = Pr(L_{A,i}|A_i)Pr(L_{A,j}|A_j)$.

Let us define I_A as the largest code rate such that the probability of a Code A decoding error goes to zero as n goes to infinity (determining I_A will be discussed in Section III-B). Now, if we choose Code A to be a long random binary code of rate $R_A < I_A$, we can safely assume that $\hat{\mathbf{A}} = \mathbf{A}$. Thus, (Step 3) we can run the APP decoder on the received channel block $\mathbf{Y}$ again, using the estimates $\hat{\mathbf{A}}$ to obtain new soft estimates $\mathbf{L}_B$. Finally, (Step 4) Code B MLD uses $\mathbf{L}_B$ to recover the remaining information bits $\hat{\mathbf{b}}$. Just as with the previous code, Code B is chosen to be a long random binary code of rate $R_B < I_B$ (determining I_B will be discussed in Section III-C), and the Code B MLD also operates under the assumption $Pr(L_{B,i}, L_{B,j}|B_i, B_j) = Pr(L_{B,i}|B_i)Pr(L_{B,j}|B_j)$.

Since the second APP decoding uses perfect *a priori* information for half of the channel symbols, the $\mathbf{L}_B$ lead to more reliable decisions than the $\mathbf{L}_A$. Therefore, we have the benefit of designing codes whose overall rate, $(R_A + R_B)/2$, may be larger than R_A, since $R_B \geq R_A$. In the next two sections, we discuss our method for determining I_A and I_B.

B. The First Code

In this section we determine achievable rates for Code A. We begin by assuming that we are using a windowed-APP detector (with window size $2w+1$) which returns log-likelihood ratios

$$L_{A,k} = \log \frac{Pr(A_k = 1 \,|\, \mathbf{Y}_{k-w}^{k+w})}{Pr(A_k = -1 \,|\, \mathbf{Y}_{k-w}^{k+w})}.$$

When the channel inputs are i.i.d. and Bernoulli one-half, we find that the random variables $L_{A,k}$ are identically distributed for k in the range $w < k \leq n - w$.

To determine the maximum achievable rate for Code A, we would need to calculate

$$\lim_{n\to\infty} \frac{1}{n} I(\mathbf{A}; \mathbf{L}_A).$$

However, this rate is much too difficult to calculate, and may not be achievable using a MLD designed for a memoryless channel.

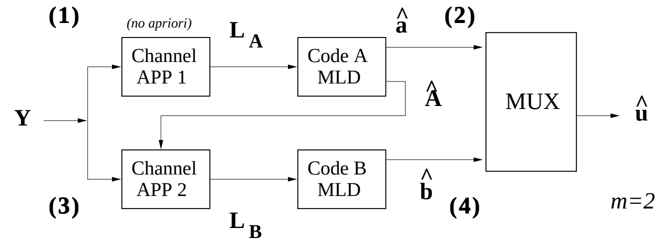

Fig. 4. Decoding structure for Fig. 3 encoder.

Instead, we focus only on the marginal density function of the detector outputs,

$$f(l \,|\, \alpha) \stackrel{def}{=} f(L_{A,k} = l \,|\, A_k = \alpha),$$

and use this to calculate the rate, $I_A = I(L_{A,k}; A_k)$. For channels described by (1) with Bernoulli one-half inputs, the marginal density will have the symmetric property, $f(l \,|\, 1) = f(-l \,|\, -1)$, reducing our calculation to the integral[2],

$$I_A = 1 - \int_{-\infty}^{\infty} f(l \,|\, 1) \log(1 + e^{-l}) dl. \quad (4)$$

Since it is not easy to solve for the density $f(l \,|\, \alpha)$, we estimate it from empirical observations. This is done by first generating a long realization of $\mathbf{Y}$, and applying an APP Detector (with no *a priori* information) to this sequence[3]. We then use a histogram of the $L_{A,i}$ values, at all times i where $A_i = 1$, to estimate $f(l \,|\, 1)$. Likewise, a histogram of all other $L_{A,i}$ values is used to estimate $f(l \,|\, -1)$.

Finally, we note that I_A is achievable with a single binary code and memoryless MLD, if in fact $Pr(L_{A,i}, L_{A,j}|A_i A_j) = Pr(L_{A,i}|A_i)Pr(L_{A,j}|A_j)$; for this to hold, we must design a system with $m \gg \nu$. However, for smaller m, we can achieve I_A by introducing a second step of interleaving. In our 2-level example, we would let Code A consist of p interleaved random codes, and have the Code A MLD correspond to p interleaved memoryless MLDs; each of the p codes and decoders operate independently of each other. For large enough p, each MLD can safely assume that $Pr(L_{A,i}, L_{A,i+kp}|A_i A_{i+kp}) = Pr(L_{A,i}|A_i)Pr(L_{A,i+kp}|A_{i+kp})$. Note in this case that, although we are using mp codes, the APP detector is employed only m times.

C. The Second Code

We proceed in a similar manner to determine I_B, the achievable rate of Code B, except in this case the APP detector has perfect information about the channel inputs from Code A (i.e., we assume $\hat{\mathbf{A}} = \mathbf{A}$). So the windowed-APP detector returns log-likelihood ratios,

$$L_{B,k} = \log \frac{Pr(B_k = 1 \,|\, \mathbf{Y}_{k-w}^{k+w}, \mathbf{A})}{Pr(B_k = -1 \,|\, \mathbf{Y}_{k-w}^{k+w}, \mathbf{A})}.$$

[2] Derived from the general formula for mutual information in [7, p. 231] using the fact that l is a log-likelihood.

[3] To avoid edge effects, we only examine samples at a distance $N \gg \nu$ away from the edges of the sequence.

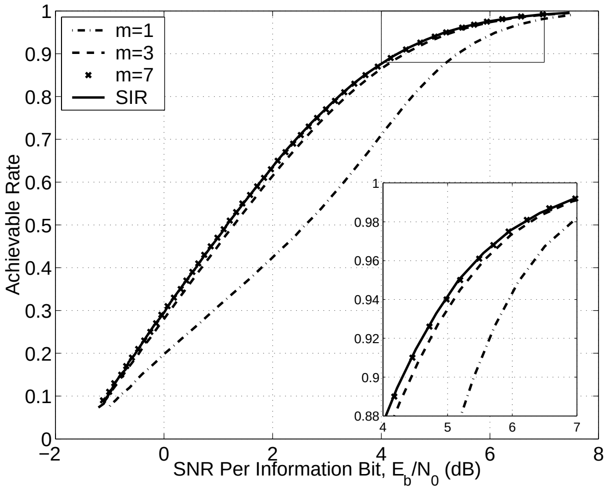

Fig. 5. Achievable rates of various m-level systems on EPR4.

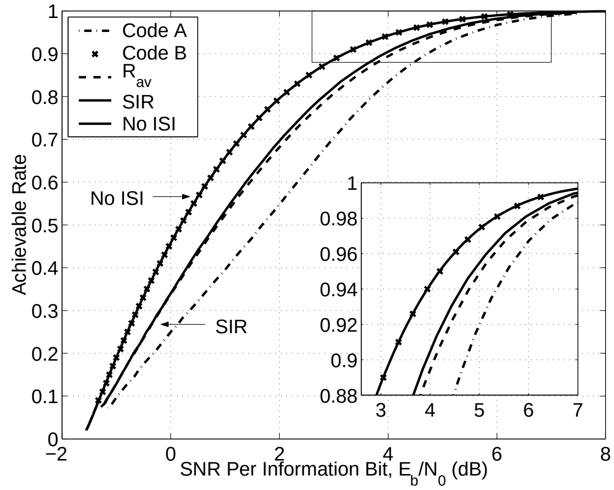

Fig. 6. Achievable rates for a 2-level system on the dicode channel.

Once again, we have random variables, $L_{B,k}$, which are identically distributed for k in the range $w < k \le n-w$, for a channel which receives i.i.d. Bernoulli one-half inputs. Moreover, the symmetric property holds also for this case, allowing us to use (4) in calculating I_B (and achieve it, by earlier arguments). The density,

$$f(l \mid \beta) \stackrel{def}{=} f(L_{B,k} = l \mid B_k = \beta),$$

is estimated by generating a second long realization of $\mathbf{Y}$, but this time the APP Detector has perfect *a priori* information for all of the Code A channel input bits and no information for the remaining Code B channel input bits. For $m > 2$, the generalization to the ith code in an interleave-by-m scheme is straightforward, since we can show that the symmetry and stationarity properties hold for each stage.

D. Simulation Results

In this section we briefly present the simulation results for the dicode and EPR4 channels. In Fig. 5, we see how closely the achievable rates of our interleaved coding system (i.e., R_{av}) approach the SIR, as estimated in the Section II-B. For example, with $m = 7$ the bounds for EPR4 are equal to the SIR (up to the plotted precision). Moreover, if we employ only $m = 3$ interleaves at rate 9/10, we lose approximately 0.2 dB relative to infinite interleaving.

In Fig. 6, we show a detailed plot of the rate distribution versus normalized SNR for the dicode channel with $m = 2$. Notice that in our finite interleaving scheme, the achievable rate of the second code matches that of a channel without ISI. In fact, for $m > \nu$, it can be proven that the mth stage of the decoder, when given perfect *a priori* information for the $m-1$ neighboring bits, reduces the ISI channel to a binary-input AWGN channel. Essentially, the perfect knowledge of $m-1$ neighboring bits makes the APP detector equivalent to subtracting all ISI due to these bits and then applying a matched filter detector.

IV. Conclusion

We present two simple Monte Carlo methods for estimating the achievable information rates of indecomposable FS channels, and provide simulation results for linear filter channels with AWGN. We use the first method to estimate the SIR of each channel and optimize the input distribution to obtain tighter bounds on capacity for the dicode channel. The second method gives an explicit code construction, consisting of interleaved random binary codes and multistage decoding. For a practical number of interleaves, simulation results indicate a negligible loss in performance relative to the SIR.

In the future, we plan to optimize the input distributions for all of the channels, thus providing tighter bounds on capacity. We are also optimizing low-density parity-check codes, to be used as components codes in the second method, using the empirical log-likelihood densities from our rate calculations.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[2] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 868–886, May 1996.

[3] D. Arnold and H. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. IEEE Int. Conf. Commun.*, (Helsinki, Finland), June 2001. in press.

[4] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. IEEE Int. Symp. on Inform. Theory*, (Washington, DC, USA), p. 283, June 2001.

[5] S. Shamai, L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1527–1539, Nov. 1991.

[6] W. Hirt, *Capacity and Information Rates of Discrete-Time Channels with Memory*. PhD thesis, E.T.H., Zurich, Switzerland, 1988.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[8] A. R. Barron, "The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem," *Ann. Probab.*, vol. 13, pp. 1292–1303, Nov. 1985.

[9] X. Chen, "Limit theorems for functionals of ergodic Markov chains with general state space," *Memoirs of the AMS*, vol. 139, May 1999.

[10] T. Richardson, M. A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.