Determining and Approaching Achievable Rates of Binary Intersymbol Interference Channels Using Multistage Decoding

Joseph B. Soriaga, *Member, IEEE*, Henry D. Pfister, *Member, IEEE*, and Paul H. Siegel, *Fellow, IEEE*

***Abstract*—By examining the achievable rates of a multistage decoding system on stationary ergodic channels, we derive lower bounds on the mutual information rate corresponding to independent and uniformly distributed (i.u.d.) inputs, also referred to as the i.u.d. information rate. For binary intersymbol interference (ISI) channels, we show that these bounds become tight as the number of decoding stages increases. Our analysis, which focuses on the marginal conditional output densities at each stage of decoding, provides an information rate corresponding to each stage. These rates underlie the design of multilevel coding schemes, based upon low-density parity-check (LDPC) codes and message passing, that in combination with multistage decoding approach the i.u.d. information rate for binary ISI channels. We give example constructions for channel models that have been commonly used in magnetic recording. These examples demonstrate that the technique is very effective even for a small number of decoding stages.**

***Index Terms*—Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm, coset codes, density evolution, finite-state channels, information rates, intersymbol interference (ISI) channels, low-density parity-check (LDPC) codes, magnetic recording, multilevel coding, multistage decoding.**

I. Introduction

ONE of the classic problems in digital communication is to reliably transmit information over the binary-input intersymbol interference channel (binary ISI channel, for short). For this channel, if $\{X_t\}$ is a binary (± 1) input sequence, then each channel output Y_t is expressible as

$$Y_t = \sum_{i=0}^{\nu} h_i X_{t-i} + N_t \tag{1}$$

where $\{N_t\}$ is modeled as additive white Gaussian noise (AWGN) with zero mean and a known variance σ^2, and $\{h_i\}_{i=0}^{\nu}$ is the known impulse response of the channel. This channel has been used as a simplified model for digital communication and data storage systems [1]. This channel is also a widely used example of a finite-state channel, since the output Y_t depends only on the current input X_t and the channel state $Q_t = (X_{t-1}, \ldots, X_{t-\nu})$. Unlike general finite-state channels, however, the channel state here is completely determined by a finite set of past inputs.

Recently, tight bounds on the capacity of binary ISI channels have been determined through the use of simulation-based estimation of achievable information rates [2]–[5]. In addition, dramatic performance improvements on coded binary ISI channels have been achieved by application of turbo-equalization techniques [6]. This paper presents bounds and constructions that complement both of these developments.

We begin by considering a multilevel code (MLC) design and a corresponding multistage decoding (MSD) algorithm for binary ISI channels. This system closely resembles the scheme proposed by Imai and Hirakawa [7]. The MLC design interleaves several independent codewords which are then successively decoded, with each decoded interleave being used toward the decoding of subsequent interleaves. By analyzing the conditional output densities at each stage of the multistage decoder, under the assumption of independent and uniformly distributed (i.u.d.) binary inputs, we determine lower bounds on the mutual information rate of the channel, which we refer to as the i.u.d. information rate, or $I_{\text{i.u.d.}}$. These lower bounds become tight as the number of stages increases to infinity. Moreover, our method of determining achievable rates focuses on marginal densities. After applying the well-known design technique for low-density parity-check (LDPC) codes given in [8] to such densities, we are able to construct explicit coding systems that closely approach $I_{\text{i.u.d.}}$.

A. Review of Achievable Rates on Binary ISI Channels

Given a stationary and ergodic input process $\{X_t\}$ and the corresponding channel output process $\{Y_t\}$, the mutual information rate for the channel, defined as

$$I(\mathcal{X};\mathcal{Y}) = \lim_{N\to\infty} \frac{1}{N} E \log_2 \frac{p\left(\boldsymbol{Y}_1^N, \boldsymbol{X}_1^N\right)}{p\left(\boldsymbol{Y}_1^N\right) p\left(\boldsymbol{X}_1^N\right)}, \tag{2}$$

determines the rate (in bits per channel use) at which information can be reliably transmitted [9], [10]. The capacity of the

Manuscript received June 2, 2004; revised December 16, 2006. This work was supported by the Center for Magnetic Recording Research at the University of California, San Diego, the Information Storage Industry Consortium (formerly National Storage Industry Consortium), and the National Science Foundation under Grants NCR-9612802 and CCR-0219582. The material in this paper was presented in part at IEEE Globecom, San Antonio, TX, November 2001 and at the Bob McEliece 60th Birthday Workshop, Pasadena, CA, May 2002.

J. B. Soriaga was with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla. He is now at Qualcomm, Inc., San Diego, CA 92121 USA (e-mail: jsoriaga@qualcomm.com).

H. D. Pfister is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: hpfister@tamu.edu).

P. H. Siegel is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla CA 92093-0401 USA (e-mail: psiegel@ucsd.edu).

Communicated by M. P. C. Fossorier, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2007.892778

channel is the supremum of $I(\mathcal{X};\mathcal{Y})$ over all input distributions, and when $\{X_t\}$ is i.u.d., we refer to the corresponding $I(\mathcal{X};\mathcal{Y})$ as $I_{\rm i.u.d.}$.

No solutions currently exist which permit an exact calculation of the mutual information rate for binary ISI channels. Early work by Hirt [11] focused on the numerical evaluation of $I_{\rm i.u.d.}$, and Shamai *et al.* [12] later bounded these results with expressions that were significantly easier to evaluate. More recently, Arnold and Loeliger [2], [5], Pfister, Soriaga, and Siegel [3], and Sharma and Singh [4] independently devised a simple Monte Carlo technique for estimating improved lower bounds on capacity, based on the Shannon–McMillan–Breiman theorem [9], [10] in combination with the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [13]. Kavčić [14] later built upon these results with an elegant input optimization procedure. Additionally, improved upper bounds using related simulation-based methods were presented by Vontobel and Arnold [15], and Yang, *et al.* [16]. These results suggest that the best lower bounds originally presented in [2]–[4] are almost equal to capacity in some cases.

Although we are ultimately concerned with achieving capacity, in this paper we focus exclusively on $I_{\rm i.u.d.}$. One important reason is that coset codes can achieve $I_{\rm i.u.d.}$ [17, p. 206], and so we seek to realize this potential by using LDPC codes with message-passing and an MLC/MSD system. Another more practical reason is that, based on the results in [2], [3], [14], the gap between $I_{\rm i.u.d.}$ and the best known lower bound on capacity is small when $I_{\rm i.u.d.}$ exceeds 0.8. Such operating regions are of prime interest to magnetic recording, and may also be of interest in optical fiber communication and optical storage.

B. Previous Coding Techniques for ISI Channels

We note that many coding systems for ISI channels have already been shown to operate reliably near $I_{\rm i.u.d.}$. Simulation results for systems using turbo-like coding architectures were given by Doulliard *et al.* [6], Ryan *et al.* [18], Souvignier *et al.* [19], and Öberg and Siegel [20], while Fan *et al.* [21] investigated architectures that incorporate LDPC codes. In these systems, the discrete-time filter in (1) is viewed as an inner code for an AWGN channel, and in a process that has come to be known as *turbo-equalization*, soft decisions are iteratively passed between the inner decoder (i.e., channel detector) and outer decoder. The asymptotic performance of turbo-equalization has been studied by several authors, including Tüchler *et al.* [22], Narayanan [23], Doan and Narayanan [24], and Thangaraj *et al.* [25]. Their methods are related to the EXIT chart analysis introduced by ten Brink [26], and, like the latter, they were also used as a guide in system design. Meanwhile, Kavčić *et al.* [27] and Varnica and Kavčić [28], [29] considered the asymptotic performance of LDPC codes by extending the analysis of Richardson *et al.* [30], [8]. In each of these cases, the threshold of turbo-equalized systems was found to be close to $I_{\rm i.u.d.}$. Such studies on the asymptotic performance of iterative decoding have predicted thresholds that are close to $I_{\rm i.u.d.}$, but do not achieve it.

Several authors have proposed the application of MLC/MSD to binary ISI channels. For instance, Filho *et al.* [31] considered an MLC/MSD architecture which uses Viterbi detection, and they designed a signal-set partitioning to increase intra-subset distances at the cost of lower rate. Kuznetsov [32] provided an alternative structured set-partitioning, and considered various other detection methods in MSD, including the soft-output Viterbi algorithm (SOVA), list-Viterbi algorithm, and the *a posteriori* probability (APP) detector. Miller and Wolf [33] examined the set of error events at each stage (for precoded channels), and accordingly designed Reed–Solomon codes at each level, while Ma *et al.* [34] combined an inner trellis code with an outer MLC/MSD system that uses LDPC component codes.

In view of this past research, the novel contribution of this paper is its information-theoretic analysis of achievable rates of MLC/MSD systems and its complementary design of LDPC codes that approach these rates arbitrarily closely. Our approach provides an alternative method for calculating the i.u.d. information rate as well as a new, practical design methodology for coding systems that closely approach this rate at any signal-to-noise ratio (SNR). The flexibility and efficiency of our coding scheme is matched only by that proposed in [28].

This paper can be viewed as a comprehensive exposition of the work reported in [3], [35], including many unpublished details and results. We note that, after the appearance of [3], [35], Narayanan and Nangare [36] revisited the MLC/MSD architecture, showing that the system complexity can be reduced, with a loss in overall rate, if pilot symbols are incorporated. Most recently, Li and Collins [37], [38] proposed natural generalizations of the successive decoding approach to a larger class of channels with memory.

The MLC/MSD scheme presented here does not use iteration between the channel detector and the component (LDPC) code at each individual level. Thus, the achievable rate for the first level of our system is similar to the "BCJR-Once Rate" introduced independently by Kavčić *et al.* [27], and our analysis of achievable rates for subsequent levels can be thought of as an extension of that earlier work. On the other hand, Ma *et al.* [34] proposed an MLC/MSD architecture that incorporates iterative decoding at each level, as part of their so-called matched information rate coding scheme.

Finally, we note that several authors have proposed spectrally efficient communication schemes for memoryless AWGN channels based upon MLC/MSD architectures with optimized component LDPC and higher order signal alphabets; see, for example, Wachsmann *et al.* [39] and Hou *et al.* [40]. However, these schemes and their analyses do not readily extend to channels with memory.

C. Overview

In Section II, we formalize MLC and MSD, along with the concept of window-APP detection (see also [27]) which simplifies the analysis of our MLC/MSD system. Section III deals with the determination of achievable information rates of MLC/MSD schemes on general stationary ergodic channels, beginning with the concept of equivalent subchannels. We develop achievable rate expressions based on the marginal conditional output densities at each stage in Section III-A, and then present in Section III-B a Monte Carlo method for easy numerical evaluation of these rates. The achievability results in this section make use of a coding theorem for "mismatched" decoders, presented in

Appendix I. In Section III-C, we prove that, for binary ISI channels, the overall system rate converges to $I_{\text{i.u.d.}}$ as the number of levels/stages $m \to \infty$.

These analytical results serve as the basis for a code design methodology, described in Section IV, that uses marginal output densities to find optimal degree distributions for irregular LDPC codes that can approach the achievable rate at each stage. The code optimization involves the application of density evolution techniques [30], [8], suitably adapted for the equivalent subchannels induced by MLC/MSD on binary ISI channels. Justification of the method is provided by an appropriate extension of the concentration theorem [30] and the verification of symmetry properties of conditional output densities for equivalent subchannels.

In Section V, we consider as specific examples the dicode and EPR4 channels, historically used as channel models in magnetic recording, with impulse responses $h(D) = \sum_{i=0}^{\nu} h_i D^i = (1-D)/\sqrt{2}$ and $h(D) = (1+D-D^2-D^3)/2$, respectively. We observe that with just $m=2$ levels on the dicode channel and $m=3$ levels on the EPR4 channel, operation within 0.1–0.2 dB of $I_{\text{i.u.d.}}$ is theoretically possible. For component codeword block lengths of 10^6, we construct a two-level system for the dicode channel that exhibits a gap to $I_{\text{i.u.d.}}$ of less than 0.3 dB at a bit-error rate (BER) of 10^{-5}, as well as a three-level system for the EPR4 channel that leaves a gap of less than 0.2 dB.

D. Notation

Random variables are denoted with upper case letters and their realizations with lower case. Vectors are represented in bold text, e.g., $\boldsymbol{X}^N = (X_1, \ldots, X_N)$, and may be further indexed explicitly, e.g., $\boldsymbol{X}_i^j = (X_i, \ldots, X_j)$. We refer to interleaved subsequences of $\boldsymbol{X}_a^b$ with the notation

$$\boldsymbol{X}_a^b(i:m) = \{X_t : a \le t \le b, t \equiv i \bmod m\}.$$

The (i)th vector in a set of length-N vectors is denoted by $\boldsymbol{X}_N^{(i)} = (X_1^{(i)}, \ldots, X_N^{(i)})$.

We use the notation $p(x)$ to abbreviate the probability $P(X=x)$, and likewise for conditional and joint probabilities of both scalar and vector random variables, e.g.,

$$p(x|q, \boldsymbol{y}_1^4) = P(X=x|Q=q, \boldsymbol{Y}_1^4 = \boldsymbol{y}_1^4).$$

Moreover, we will use $p(\cdot)$ when denoting both probability mass functions and probability density functions. For instance, if Y takes values in a continuous alphabet, we write $p(y)$ for the probability density function $f_Y(y)$. The meaning should be easily deduced from context, although more explicit notation will sometimes be provided for clarity. Regarding entropy, we use the general definition $H(X) = -E \log_2 p(X)$, remembering that the function $p(x)$ was implicitly defined to be $P(X=x)$. We use this form as it can be meaningfully interpreted for both discrete and continuous random variables. Mutual information is similarly expressed as $I(X;Y) = E \log_2(p(X,Y)/p(X)p(Y))$.

II. Multilevel Encoding and Multistage Decoding

In this section, we formulate the basic structure of MLC/MSD considered in this paper, detailing the multilevel encoder, the stage-by-stage operation for decoding, and the APP detector used at each of these stages. Similar to Kavčić *et al.* [27], we use the concept of a window-APP detector to simplify analysis. Our system essentially resembles that of Imai and Hirakawa [7], except that here we directly transmit each binary m-tuple from the m-level coding system rather than map it to a higher order constellation, and here we employ at each decoding stage a window-APP detector which operates on overlapping blocks of symbols, in order to account for channel memory.

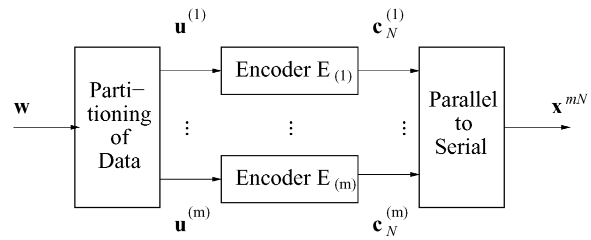

Fig. 1. Multilevel encoder.

A. Multilevel Encoder

For an m-level encoder (see Fig. 1), a block of data bits $\boldsymbol{w}$ is partitioned into m subblocks of varying size, and each of these information blocks is separately mapped to a codeword of length N. For the ith code of rate $R_m^{(i)}$, its codeword is denoted by $\boldsymbol{c}_N^{(i)} = (c_1^{(i)}, \ldots, c_N^{(i)})$. All of these m codewords are then interleaved into

$$\boldsymbol{x}^{mN} = \left(c_1^{(1)}, \ldots, c_1^{(m)}, c_2^{(1)}, \ldots, c_2^{(m)}, \ldots, c_N^{(1)}, \cdots, c_N^{(m)}\right) \tag{3}$$

which is transmitted through the channel. This interleaving can also be written element-wise with $x_{i+(j-1)m} = c_j^{(i)}$, or more concisely with $\boldsymbol{x}^{mN}(i:m) = \boldsymbol{c}_N^{(i)}$. Sometimes we use $\kappa(j)$ to compactly denote $i + (j-1)m$, whenever the values of i and m are understood.

This m-level encoder can be viewed as a single encoder for an $(n,k) = (mN, NR_m^{(1)} + \ldots + NR_m^{(m)})$ binary block code, with an overall rate $R_{\text{av},m} = \sum_{i=1}^m R_m^{(i)}/m$. The structure of each component code is discussed in Section IV.

B. Multistage Decoder for Binary ISI Channels

Given a received vector $\boldsymbol{y}^{mN}$, the m interleaved codewords are recovered with MSD (Fig. 2) as follows, proceeding from stage $i=1$ to m:

1) *Channel block-APP detection.* At stage i, the channel detector determines the vector of log-APP ratios $\boldsymbol{l}_N^{(i)}$, where each element $l_j^{(i)} = \log(p_j^{(i)}/(1-p_j^{(i)}))$ and

$$p_j^{(i)} = P(X_{\kappa(j)} = 1 | \boldsymbol{Y}^{mN} = \boldsymbol{y}^{mN}, \boldsymbol{X}^{mN}(1:m) = \hat{\boldsymbol{c}}^{(1)}, \vdots, \boldsymbol{X}^{mN}(i-1:m) = \hat{\boldsymbol{c}}^{(i-1)})$$

is the APP calculated assuming the $\{X_t\}$ are i.u.d. Each $\hat{\boldsymbol{c}}_N^{(i')}$ is a codeword decision from the previous interleaved level i'. (For the first stage, the conditioning on X_t is absent from the APP expression.) The log-APP ratios $l_j^{(i)}$ for all $j = 1, \ldots, N$ can be calculated using the BCJR algorithm

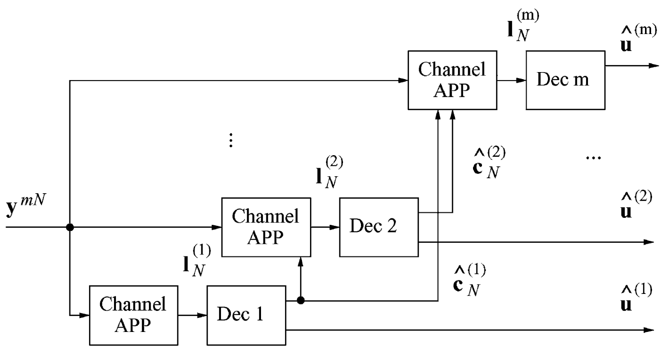

Fig. 2. Multistage decoder.

[13] on the entire received sequence $\boldsymbol{y}^{mN}$, keeping only every mth estimate starting from time i.

2) *Codeword recovery*. The ith component decoder determines $\hat{\boldsymbol{c}}^{(i)}$ from $\boldsymbol{l}_N^{(i)}$.

To lower the computational complexity, the channel detector and component decoder at each stage are separated, and they do not iteratively exchange information.

C. Window-APP Detection

We now describe the *window-APP detector* used in our analysis of MLC/MSD.

Definition 1: For MSD on binary ISI channels, the *window-APP* at time index j is the APP based on a fixed window of observations with *side length* w. Formally, at stage i we calculate for each $j = 1, \dots, N$,

$$\begin{aligned} p_j^{(i)} = P(X_{\kappa(j)} = 1 | &\boldsymbol{Y}_a^b = \boldsymbol{y}_a^b, \\ &\boldsymbol{X}_a^b(1{:}m) = \hat{\boldsymbol{x}}_a^b(1{:}m), \\ &\vdots \\ &\boldsymbol{X}_a^b(i-1{:}m) = \hat{\boldsymbol{x}}_a^b(i-1{:}m)) \\ a = \max(1, \kappa(j) - w) \quad & b = \min(mN, \kappa(j) + w) \end{aligned} \tag{4}$$

where we let $\hat{\boldsymbol{x}}^{mN}(i'{:}m) = \hat{\boldsymbol{c}}_N^{(i')}$ for each previous interleave i'. (Again, the conditioning on X_t is absent from the APP expression for the first stage.) The log-APP ratios are accordingly defined as $l_j^{(i)} = \log(p_j^{(i)}/(1-p_j^{(i)}))$. Any index j such that $1 + w \leq \kappa(j) \leq mN - w$ is referred to as an *interior point*, and at these indices the detector has a full window of $2w + 1$ observations.

According to the definition, if the input is independent and identically distributed, then the output log-APP ratios $L_j^{(i)}$ are identically distributed for all interior points j. In (4), the definitions of the upper and lower indices of the observations in the window reflect the truncation effect at the edge points. For a fixed w, the fraction of edge points asymptotically goes to zero as the block length $N \to \infty$, so our analysis concentrates on the detector behavior at interior points.

The MSD simulations in Section IV all use block-APP detection, which is computationally more efficient than window-APP detection. For large enough w and N, the window-APP and block-APP detector differ negligibly in performance. In fact, even the suboptimal window-based message-passing detector investigated by Kurkoski *et al.*[41] achieved the performance of the block-APP detector for large enough w. (For certain partial-response channels, the window-based message-passing detector in [41] produced an error floor, caused by the existence of ambiguous input sequences, i.e., bi-infinite, nonzero input sequences which have zero output energy. However, the error floor could be made arbitrarily low by sufficiently enlarging the window size, or it could be eliminated completely with an appropriately chosen precoder.) A more detailed discussion of the asymptotic equivalence of window-APP and block-APP detection, based on the exponential forgetting arguments of Le Gland and Mevel [42], [43] (see also [44]), can be found in [45].

III. Stage-Wise Analysis of Achievable Rates for MLC/MSD

Given an m-level system, we can determine achievable information rates by separately examining each level/stage assuming previous decoding stages were successful. This is justified by the observation that, if E_{MSD} is the set of all MSD error events and E_i is the set of events in which stage i fails, then

$$\begin{aligned} P(E_{\text{MSD}}) = P(E_1) + P(\bar{E}_1 \cap E_2) + \cdots \\ + P(\bar{E}_1 \cap \cdots \cap \bar{E}_{m-1} \cap E_m). \end{aligned} \tag{5}$$

Additionally, we assume that the inputs from previous stages are i.u.d., thus reflecting the average performance when each of the component codes is taken from the Shannon ensemble (i.e., each is a codebook randomly generated from an i.u.d. source). Under these two assumptions, we refer to the resulting behavior at each level/stage as an *equivalent subchannel*.

Definition 2: The ith *equivalent subchannel* in an MLC/MSD system represents the transfer characteristics from the ith interleaved input to the output of a window-APP detector (of side length w) that is given perfect decisions for all previous interleaves. For N uses of the channel, the inputs are $\boldsymbol{c}_N^{(i)}$; the outputs are $\boldsymbol{l}_N^{(i)}$, as given in Definition 1 except that $\hat{\boldsymbol{c}}^{(i')}$ is replaced with $\boldsymbol{c}^{(i')}$ to reflect perfect *a priori* information; and all code bits $\boldsymbol{c}^{(i')}$ from other levels are assumed to be i.u.d.

This definition is illustrated in Fig. 3. Note that the concept of equivalent subchannels in an MLC/MSD system is not limited to binary ISI channels. Therefore, for the larger class of stationary ergodic channels, we present below a closed-form expression for the achievable information rate $R_m^{(i)}$ on each equivalent subchannel based on marginal conditional output densities. We remark that the rate $R_m^{(i)}$ is shown to be achievable in the sense that, in the limit of infinite codeword block length, a zero probability of decoding error can be ensured for all rates below $R_m^{(i)}$ even if the decoder assumes the equivalent subchannel is memoryless. It is then shown in Section III-C how these rate expressions can be simplified in the case of binary ISI channels, and the resulting formulas allow us to prove that the overall MLC/MSD system rate converges to $I_{\text{i.u.d.}}$ as $m \to \infty$. In Section IV, we consider designing an MLC/MSD system with LDPC component codes to approach these rates. For completeness, the general idea of coding based on partial information of the channel

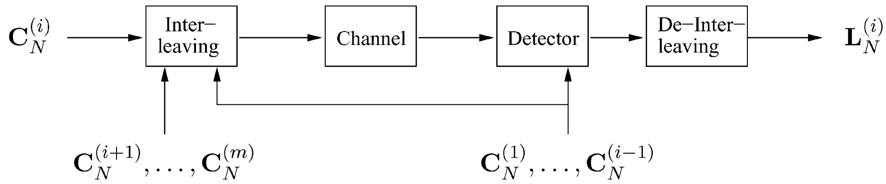

Fig. 3. *Equivalent subchannel* for level/stage i. Decisions from previous stages are assumed to be perfectly known, and i.u.d.

law (e.g., marginal densities), or possibly even mismatched information, is elaborated upon in Appendix I.

The proofs throughout this section rely on the following elementary proposition.

Proposition 1: If A, B, and Y are random variables, with A and B independent, then $I(A;Y|B) = I(A;Y,B)$.

Proof: Using the chain rule for mutual information, [9, p. 22], $I(A;Y,B) = I(A;B) + I(A;Y|B) = I(A;Y|B)$, since $I(A;B) = 0$. □

A. Achievable Rates of Equivalent Subchannels

By examining equivalent subchannels, we obtain the following analytical results for achievable rates.

Theorem 1: Let $\{X_t\}$ be an i.u.d. binary-input sequence and let $\{Y_t\}$ be the corresponding output from a stationary ergodic channel. Consider an MLC/MSD system with m levels and window-APP detector side length w. On the first equivalent subchannel, the rate

$$R_m^{(1)} = R^{(1)} \stackrel{\text{def}}{=} I\left(X_1; \boldsymbol{Y}_{1-w}^{1+w}\right) \tag{6}$$

is an achievable information rate. For the ith equivalent subchannel, $1 < i \leq m$, the rate

$$\begin{aligned} R_m^{(i)} \stackrel{\text{def}}{=} I(X_i; \boldsymbol{Y}_{i-w}^{i+w} | \boldsymbol{X}_{i-w}^{i+w}&(1:m), \\ &\vdots \\ &\boldsymbol{X}_{i-w}^{i+w}(i-1:m)) \end{aligned} \tag{7}$$

is achievable. Moreover, for any interior point j, $I(C_j^{(i)}; L_j^{(i)}) = R_m^{(i)}$, and so $R_m^{(i)}$ is achievable even with knowledge of only the marginal density $p(c_j^{(i)}, l_j^{(i)})$.

The rightmost expression in (6) is the mutual information between an input bit and the output values considered by the window-APP detector. The last statement of the theorem motivates the design of component LDPC codes with message-passing decoders using the marginal density $p(c_j^{(i)}, l_j^{(i)})$. (See Section IV.) The rate $R_m^{(1)}$ is independent of the number of levels m because no prior decisions were given. For this reason, the notation $R^{(1)}$ is sometimes used. This rate is also equivalent to the "BCJR-Once Rate" independently reported by Kavčić [27].

Proof (Theorem 1): To prove that $R_m^{(i)}$ is achievable, even with a decoding rule based on marginal densities, we need only show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \log_2 \left(\frac{p\left(C_j^{(i)}, L_j^{(i)}\right)}{p\left(C_j^{(i)}\right) p\left(L_j^{(i)}\right)} \right) = R_m^{(i)} \quad \text{(a.s.)} \tag{8}$$

Having established (8), we can use the mismatched coding theorem in Appendix I to demonstrate that a typical-set decoder can be constructed to achieve the rate $R_m^{(i)}$ using some code from the Shannon ensemble.

We begin with the stage $i = 1$. Let j be any interior point. Recall that $\kappa(j) = i + (j-1)m$ describes the mapping from the index j in the ith interleaved codeword to its time index $\kappa(j)$ in the transmitted sequence. From Definition 1 of the window log-APP ratio, it is obvious that $L_j^{(1)}$ is a sufficient statistic for $\boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w}$ when determining $C_j^{(1)} = X_{\kappa(j)}$. Thus, $I(C_j^{(1)}; L_j^{(1)}) = I(X_{\kappa(j)}; \boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w})$ [9, p. 37]. But, by stationarity

$$I\left(X_{\kappa(j)}; \boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w}\right) = I\left(X_1; \boldsymbol{Y}_{1-w}^{1+w}\right) = R^{(1)}.$$

For any N, all but a constant number of summands (i.e., those corresponding to interior points) in (8) are identically distributed and have the expectation $I(C_j^{(1)}; L_j^{(1)}) = R^{(1)}$. Therefore, as a consequence of the ergodic theorem [46], one finds that the limit in (8) is almost surely $R^{(1)}$.

Extending this proof to cases $i > 1$ is straightforward. Here, when determining $C_j^{(i)}$ for any interior point j, the log-APP ratio $L_j^{(i)}$ can be used as a sufficient statistic for both the output window $\boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w}$ and the set of input windows, $\boldsymbol{X}_{\kappa(j)-w}^{\kappa(j)+w}(1{:}m), \ldots, \boldsymbol{X}_{\kappa(j)-w}^{\kappa(j)+w}(i-1{:}m)$, from previous interleaves. Furthermore, we have

$$\begin{aligned} I\left(C_j^{(i)}; L_j^{(i)}\right) = I\Big(X_{\kappa(j)}; \boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w}, &\boldsymbol{X}_{\kappa(j)-w}^{\kappa(j)+w}(1:m), \\ &\vdots \\ &\boldsymbol{X}_{\kappa(j)-w}^{\kappa(j)+w}(i-1{:}m)\Big) \end{aligned}$$

[9, p. 37], which in turn equals

$$\begin{aligned} I(X_{\kappa(j)}; \boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w} | \boldsymbol{X}_{\kappa(j)-w}^{\kappa(j)+w}&(1{:}m), \\ &\vdots \\ &\boldsymbol{X}_{\kappa(j)-w}^{\kappa(j)+w}(i-1:m)) \end{aligned}$$

by Proposition 1, because the input bits are independent and identically distributed. By stationarity, we have $I(C_j^{(i)}; L_j^{(i)}) = R_m^{(i)}$. Since $L_j^{(i)}$ is identically distributed for all interior points j, we conclude that, for any N, all but a constant number of summands in (8) are identically distributed with mean $R_m^{(i)}$. Applying the ergodic theorem again, we conclude that (8) holds for $i > 1$. □

B. Monte Carlo Method for Calculation of Achievable Rates

The closed-form expressions of Theorem 1 represent a multidimensional integral which may be too complex to evaluate easily. On the other hand, in the proof for Theorem 1 we find an almost sure convergence property which suggests that $R_m^{(i)}$ might also be evaluated through Monte Carlo simulation. For instance, if one can simulate the equivalent subchannel to obtain a set of observed input and output sequences, then the estimated density $p(c_j^{(i)}, l_j^{(i)})$ (i.e., from a histogram) can be used to approximate the mutual information $I(C_j^{(i)}; L_j^{(i)})$.

There is of course a looseness to this approximation due to the finite precision of the histogram. Fortunately, we can take another Monte Carlo approach, embodied in Theorem 2 below, which does not suffer from such inaccuracy. This method again involves equivalent subchannel simulation, but uses a simple function evaluation for each input–output pair $(C_j^{(i)}, L_j^{(i)})$ to form a sample average. The sample average converges to $R_m^{(i)}$ almost surely as $N \to \infty$. It should be noted that, for the case $i = 1$, this method of calculating $R^{(i)}$ is equivalent to the calculation of the "BCJR-Once Rate" in [47], though our formulation is slightly different and more readily implementable.

Theorem 2: For the ith equivalent subchannel of any MLC/MSD system, if $\boldsymbol{C}_N^{(i)}$ is a sequence of i.u.d. inputs with corresponding output sequence $\boldsymbol{L}_N^{(i)}$, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} f\left(C_j^{(i)} \cdot L_j^{(i)}\right) = R_m^{(i)} \quad \text{(a.s.)} \tag{9}$$

where $f(l) = 1 - \log_2(1 + e^{-l})$, and $R_m^{(i)}$ is defined as in Theorem 1.

Remark 1: The function $f(l)$ is identical to that used in [48] for the capacity expression of binary output-symmetric memoryless channels followed by a log-likelihood-ratio detector.

Proof (Theorem 2): Assume j is any interior point. To prove the theorem, we need only show that $E[f(C_j^{(i)} \cdot L_j^{(i)})] = R_m^{(i)}$, because the a.s. convergence in (9) will then follow from the same reasoning used to verify (8) in the proof of Theorem 1.

First let us consider $i = 1$, and define $p(c|\boldsymbol{y}_{-w}^{w}) = P(C_j^{(1)} = c|\boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w} = \boldsymbol{y}_{-w}^{w})$. The window log-APP ratio is given by $L_j^{(1)} = \log(p(1|\boldsymbol{Y}_{-w}^{w})/p(-1|\boldsymbol{Y}_{-w}^{w}))$, from which we conclude that

$$\begin{aligned} f\left(L_j^{(1)}\right) &= 1 - \log_2\left(1 + \frac{p\left(-1|\boldsymbol{Y}_{-w}^{w}\right)}{p\left(1|\boldsymbol{Y}_{-w}^{w}\right)}\right) \\ &= 1 + \log_2 p\left(1|\boldsymbol{Y}_{-w}^{w}\right) \end{aligned}$$

and, similarly, $f(-L_j^{(1)}) = 1 + \log_2 p(-1|\boldsymbol{Y}_{-w}^{w})$. Together, these imply

$$f\left(C_j^{(1)} \cdot L_j^{(1)}\right) = 1 + \log_2 p\left(C_j^{(1)} \middle| \boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w}\right).$$

Therefore, $E[f(C_j^{(1)} \cdot L_j^{(1)})] = 1 - H(C_j^{(1)}|\boldsymbol{Y}_{\kappa(j)-w}^{\kappa(j)+w}) = R_m^{(1)}$, which completes the proof for the $i = 1$ case.

As in the proof of Theorem 1, the argument for the case $i = 1$ extends easily to the cases where $i > 1$. □

C. Convergence to $I_{\text{i.u.d.}}$

The expressions for the achievable rates in Theorem 1 are applicable to arbitrary stationary ergodic channels. In the case of binary ISI channels, we can further show that the overall MLC/MSD system rate $R_{\text{av},m}$ converges to $I_{\text{i.u.d.}}$ as $m \to \infty$. Our analysis will rely upon the following lemma.

Lemma 1: Consider an ISI channel with input process $\{X_t\}$ and output process $\{Y_t\}$, and let the state of the channel at time t be $Q_t = \boldsymbol{X}_{t-\nu}^{t-1}$. Then, $I_{\text{i.u.d.}} = \lim_{N\to\infty} I(X_1; \boldsymbol{Y}_1^N|Q_1)$.

Proof: By applying the chain rule for mutual information [9, p. 22]

$$\begin{aligned} I_{\text{i.u.d.}} &= \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{X}^N; \boldsymbol{Y}^N) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} I(X_{N-i+1}; \boldsymbol{Y}^N|\boldsymbol{X}^{N-i}). \end{aligned} \tag{10}$$

When $(N-i) \geq \nu$, the state of the channel at time $(N-i+1)$ is known, and we have

$$I(X_{N-i+1}; \boldsymbol{Y}^N|\boldsymbol{X}^{N-i}) = I(X_{N-i+1}; \boldsymbol{Y}^N|Q_{N-i+1}, \boldsymbol{X}^{N-i-\nu})$$

which equals

$$\begin{aligned} &I\left(X_{N-i+1}; \boldsymbol{Y}_{N-i+1}^{N} \middle| Q_{N-i+1}, \boldsymbol{X}^{N-i-\nu}\right) \\ &\quad + I\left(X_{N-i+1}; \boldsymbol{Y}^{N-i} \middle| Q_{N-i+1}, \boldsymbol{X}^{N-i-\nu}, \boldsymbol{Y}_{N-i+1}^{N}\right) \end{aligned}$$

by the chain rule. From the underlying Markov dependence, $\boldsymbol{Y}^{N-i}$ is independent of any terms in the future (i.e., X_{N-i+1}) once it is conditioned on Q_{N-i+1}, implying that the second term above is 0. Similarly, X_{N-i+1} and $\boldsymbol{Y}_{N-i+1}^{N}$ are independent of the past when conditioned on Q_{N-i+1}, so the first term reduces to

$$I\left(X_{N-i+1}; \boldsymbol{Y}_{N-i+1}^{N} | Q_{N-i+1}\right).$$

This last expression equals $I(X_1; \boldsymbol{Y}^i|Q_1)$ by stationarity (shift backward by $N - i$), and after substituting this into (10) and ignoring the first ν summands, we get

$$I_{\text{i.u.d.}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N-\nu} I(X_1; \boldsymbol{Y}^i|Q_1). \tag{11}$$

Notice that the sequence $I(X_1; \boldsymbol{Y}^i|Q_1)$ is monotonically nondecreasing in i and bounded from above by 1. Therefore, its limit $\lim_{i\to\infty} I(X_1; \boldsymbol{Y}^i|Q_1)$ exists and equals the limit of the Cesáro mean (e.g., [9, p. 64]) in (11). This proves the lemma. □

Now, we observe that for $i > \nu$, perfect decisions accumulate to yield perfect state knowledge periodically throughout the block. Such state information results in component rates $R_m^{(i)}$ which actually exceed $I_{\text{i.u.d.}}$. This property can be used to prove convergence of the system rate to $I_{\text{i.u.d.}}$.

Theorem 3: If i exceeds the channel memory ν, and the window side length satisfies $w \geq m + \nu - i$, then on the ith equivalent subchannel

$$R_m^{(i)} = I(X_1; \boldsymbol{Y}_1^{m-i+\nu+1}|Q_1, Q_{m-i+\nu+2}) \geq I_{\text{i.u.d.}}. \tag{12}$$

(Recall that the channel states are given by $Q_1 = \boldsymbol{X}_{1-\nu}^{0}$ and $Q_{m-i+\nu+2} = \boldsymbol{X}_{m-i+2}^{m-i+\nu+1}$.) Consequently, the achievable system rate of MLC/MSD converges to $I_{\text{i.u.d.}}$ as

$$|I_{\text{i.u.d.}} - R_{\text{av},m}| = O(m^{-1}).$$

Proof: When $i > \nu$ and $w \geq m + \nu - i$, the channel inputs upon which the mutual information in (7) is conditioned, include $X_1, \ldots, X_{i-1}$, as well as $X_{1+m}, \ldots, X_{i-1+m}$. More specifically, these inputs determine states Q_i and $Q_{m+\nu+1}$, so (7) reduces to

$$R_m^{(i)} = I\left(X_i; \boldsymbol{Y}_i^{m+\nu} | Q_i, Q_{m+\nu+1}\right),$$

because any inputs or outputs that come before or after these two states are conditionally independent of X_i. By stationarity, this is equal to the mutual information in (12).

Now consider the following inequality:

$$I(X_1; \boldsymbol{Y}^N | Q_1) \leq I(X_1; \boldsymbol{Y}^N, Q_{m+\nu-i+2} | Q_1).$$

We can combine this with the fact that

$$I(X_1; \boldsymbol{Y}^N, Q_{m+\nu-i+2} | Q_1) = I(X_1; \boldsymbol{Y}^N | Q_1, Q_{m+\nu-i+2})$$

which holds because X_1 and $Q_{m+\nu-i+2}$ are independent (Proposition 1). For all $N > m + \nu - i + 1$, the right-hand side can be reduced to $I(X_1; \boldsymbol{Y}_1^{m+\nu-i+1} | Q_1, Q_{m+\nu-i+2})$ because of the underlying Markov dependence, and this in turn equals $R_m^{(i)}$. Hence, for all sufficiently large N, we have $I(X_1; \boldsymbol{Y}^N | Q_1) \leq R_m^{(i)}$, which by Lemma 1 proves (12).

Finally, (12) allows us to bound the gap in rate

$$I_{\text{i.u.d.}} - \overbrace{\frac{1}{m}\sum_{i=1}^{m} R_m^{(i)}}^{R_{\text{av},m}} \leq I_{\text{i.u.d.}} - \frac{1}{m}\sum_{i=1}^{\nu} R_m^{(i)} - \frac{1}{m}\sum_{i=\nu+1}^{m} I_{\text{i.u.d.}}$$
$$= \frac{1}{m}\sum_{i=1}^{\nu}(I_{\text{i.u.d.}} - R_m^{(i)}) \leq \frac{\nu}{m}.$$

By the data processing inequality, this difference can never be negative, and so the proof is complete. □

We remark that an interesting property arises when $m > \nu$. Here, the mth stage can cancel all of the ISI, and thus the mth equivalent subchannel is reduced to a binary-input AWGN channel.

Corollary 1: When $m > \nu$, the mth equivalent subchannel (with window side length $w \geq \nu$) can be reduced to a binary-input AWGN channel with SNR $\sum_{i=0}^{\nu} h_i^2/\sigma^2$. Hence, $R_m^{(m)}$ is equal to the capacity of this channel.

Proof: See Appendix II. □

IV. Approaching Achievable Rates of MLC/MSD With LDPC Codes

For binary-input memoryless channels that are symmetric, it is well known from Richardson *et al.* [30], [8], that LDPC codes can be designed to closely approach the channel capacity. Their method relies on applying an analysis technique, referred to as *density evolution*, to the conditional output density of the channel. Here, we show that if the techniques in [30], [8] are applied to the marginal conditional output densities at each level/stage of MLC/MSD, then component LDPC codes (with an additional coset selector) can be designed to approach the rate $R_m^{(i)}$ defined in Section III. Ultimately, then, MLC/MSD systems can be constructed to approach $I_{\text{i.u.d.}}$ for binary ISI channels.

A. LDPC Codes and Their Cosets

Recall that a rate $K{:}N$ binary linear code $\mathcal{C}_N$ is a linear subspace of $\mathbb{F}_2^N$, consisting of 2^K vectors, where $\mathbb{F}_2 = \{0, 1\}$ is the binary field. The code $\mathcal{C}_N$ can be described in terms of a $K \times N$ generator matrix G as the set $\mathcal{C}_N = \{\boldsymbol{u}G{:}\boldsymbol{u} \in \mathbb{F}_2^K\}$, or, equivalently, in terms of an $(N-K) \times N$ parity-check matrix H as the set $\mathcal{C}_N = \{\boldsymbol{c}{:}\boldsymbol{c} \in \mathbb{F}_2^N, \boldsymbol{c}H^T = 0\}$.

A coset of a linear code (coset code, for short) is the translated set of vectors, $\mathcal{C}_N \oplus \boldsymbol{s}$, where $\boldsymbol{s} \in \mathbb{F}_2^N$, with "$\oplus$" being the addition operation. The vector $\boldsymbol{s}$ is referred to as the coset selector.

Coset codes can readily be incorporated into an MLC/MSD system. Each code bit x is modulated into the value $(-1)^x$ prior to transmission. For each level i, the partitioned data block $\boldsymbol{U}_K^{(i)}$ is encoded into $\boldsymbol{C}_N^{(i)} = \boldsymbol{U}_N^{(i)}G^{(i)} \oplus \boldsymbol{s}_N^{(i)} = \tilde{\boldsymbol{C}}_N^{(i)} \oplus \boldsymbol{s}_N^{(i)}$, where the matrix $G^{(i)}$ defines the ith component linear code and $\boldsymbol{s}_N^{(i)}$ defines the coset. At the ith stage of MSD, we adjust the sign of the log-APP ratio prior to the decoder for the ith linear code, as prescribed by the coset vector $\boldsymbol{s}_N^{(i)}$. That is, we determine $\tilde{\boldsymbol{L}}_N^{(i)} = (\tilde{L}_1^{(i)}, \ldots, \tilde{L}_N^{(i)})$, where

$$\tilde{L}_j^{(i)} = \begin{cases} L_j^{(i)}, & \text{if } s_j^{(i)} = 0 \\ -L_j^{(i)}, & \text{otherwise} \end{cases}$$

and $L_j^{(i)}$ is the output log-APP ratio from the detector.

The ensemble of irregular LDPC codes we consider is the well-known ensemble introduced in [30, p. 601]. Here, the bipartite graph representation of the parity-check matrix is characterized by a block length N, a fraction λ_i (respectively, ρ_i) of edges connected to bit (respectively, check) nodes of degree i, and a distribution of bit and check degrees $\lambda(x) = \sum_2^{\lambda_{\max}} \lambda_i x^{i-1}$ and $\rho(x) = \sum_{i=2}^{\rho_{\max}} \rho_i x^{i-1}$, respectively.

B. Concentration, Density Evolution, and Code Design

A key result in the analysis of LDPC codes under message-passing decoding is the concentration theorem [30, p. 613], which shows that the performance of randomly chosen graphs from $\mathcal{C}_N(\lambda, \rho)$ is tightly concentrated around that of a cycle-free local decoding graph for sufficiently large block lengths. The application of code design techniques from [30], [8] to our case requires an extension of the concentration theorem.

This extension can be derived using results in [27] and [49], [40]. First, Kavčić *et al.* [27] develop a generalization of the concentration theorem for the case of message passing on the joint LDPC-channel graph. Their results can be readily modified to address each decoding stage of MLC/MSD. Second, Hou *et al.* [49], [40], extend the concentration theorem to MLC/MSD with component LDPC codes on an AWGN channel. Building upon these earlier works, and taking into account the fact that the equivalent subchannels in our system are not memoryless, we can prove the following concentration theorem for MLC/MSD systems on binary ISI channels.

Theorem 4: Let $\mathcal{N}_e^{2\ell}$ be the ℓ-iteration local decoding neighborhood for a bit-to-check edge e, as defined in [30, p. 602]. This neighborhood is said to be tree-like if all bit and check nodes in $\mathcal{N}_e^{2\ell}$ are distinct, and bit nodes receive messages from the channel which are independent when conditioned upon the input sequence.

For a graph chosen from the random ensemble $\mathcal{C}_N(x^{d_v-1}, x^{d_c-1})$, a randomly chosen coset selector $\boldsymbol{s}_N^{(i)}$, and a realization of the ith equivalent subchannel with an all-zero codeword input, let Z be the number of incorrect bit-to-check messages at the ℓth iteration of message-passing decoding, and let p_ℓ^* be the expected fraction of incorrect messages passed along an edge e with a tree-like neighborhood $\mathcal{N}_e^{2\ell}$ at the ℓth iteration. For any $\epsilon > 0$ and $N > 2\gamma/\epsilon$

$$P\left(|Z - d_v N p_\ell^*| > d_v N \epsilon\right) \le 2e^{-\beta\epsilon^2 N} \tag{13}$$

where $\beta = \beta(d_v, d_c, \ell)$ and $\gamma = \gamma(d_v, d_c, \ell)$.

For the sake of brevity and because the modifications to [27], [40] are straightforward, we do not prove this theorem here. A detailed proof can be found in [45]. It should be noted that we have modified the definition of tree-like neighborhoods slightly in view of the fact that the equivalent subchannel has memory. However, the output log-APP ratios are conditionally independent if they are sufficiently separated in time. This follows from the channel definition in (1) and the properties of the window-APP detector. We also note that (13) is an average over all possible coset selectors, which implies that the theorem also holds for codewords other than the all-zero codeword. Hence, as $N \to \infty$, there is a vanishing probability of simultaneously choosing a code graph, coset selector, codeword, and noise realization such that the fraction of erroneous messages (among all edges) deviates from p_ℓ^* by more than ϵ. If we denote by $p_\ell^*(i)$ the value p_ℓ^* for the ith randomized equivalent subchannel, then the overall performance across all equivalent subchannels for this MLC/MSD system after ℓ iterations should approach $(1/m)\sum_{i=1}^m p_\ell^*(i)$ as $N \to \infty$.

Finally, Lemma 2 below shows that the conditional output densities of equivalent subchannels satisfy two symmetry properties. These properties are also shared by the binary-input memoryless channels considered in [8], and more importantly, allow for direct application of density evolution and the stability criterion of [8].

Lemma 2: For the ith equivalent subchannel and any interior point j, the conditional output density

$$p_{(i)}(l|c) = P(L_j^{(i)} = l | C_j^{(i)} = c)$$

satisfies *output-symmetry*, i.e., $p_{(i)}(l|c) = p_{(i)}(-l|-c)$, and *exponential-symmetry*, i.e., $p_{(i)}(l|c) = e^l p_{(i)}(-l|c)$.

The proof is straightforward once one notes that for binary ISI channels, $p(y_k|\boldsymbol{x}_{k-\nu}^k) = p(-y_k|-\boldsymbol{x}_{k-\nu}^k)$, and when $\{X_t\}$ is i.u.d., this leads to $p(\boldsymbol{y}|\boldsymbol{x}_{k-w-\nu}^{k+w}) = p(-\boldsymbol{y}|-\boldsymbol{x}_{k-w-\nu}^{k+w})$. Full details can be found in [45].

V. Design Examples and Simulation Results

We now demonstrate this design methodology on some binary ISI channel models relevant to magnetic recording. For reference, $I_{\text{i.u.d.}}$ and capacity lower bounds are computed using techniques from [2], [3], and all information rates and simulation results are plotted versus the SNR, $E_s/N_0 = 1/(2\sigma^2)$. As discussed earlier, in all simulations we employ a block-APP detector for efficiency, rather than the window-APP detector used in our analysis. For the component code design, the ensembles $\mathcal{C}_N(\lambda, \rho)$ were optimized using approximate linear programming methods such as [48], [50].

Example 1: Dicode channel, $h(D) = (1-D)/\sqrt{2}$.

The first step of MLC/MSD system design is to determine the number of levels to be used. Following Section III-B, we calculate the achievable rates for various m-level systems and plot them in Fig. 4(a). (Unless otherwise specified, each point in the curve is based on a Monte Carlo simulation of 10^7 samples.) Immediately, we see that $I_{\text{i.u.d.}}$ can be closely approached with small m. For example, the figure insert shows that, at an overall system rate of 0.9 bit/channel-use, 1-level, 2-level, and 5-level coding schemes with MSD can operate within 0.79, 0.16, and 0.01 dB of $I_{\text{i.u.d.}}$, respectively. Similarly, Table I shows that $R_{\text{av},m}$ comes within 0.01-bit/channel use of $I_{\text{i.u.d.}}$ for m as small as 2 (for this table, the Monte Carlo simulation used 10^6 samples). Therefore, we can use $m = 2$ levels with negligible performance loss. In Fig. 4(b), we illustrate the distribution of coding among the two levels by plotting the component achievable rates. Notice that $R_2^{(2)} \ge I_{\text{i.u.d.}}$, in accordance with Theorem 3.

Next, we design component LDPC codes for a two-level system using the methods in Section IV. That is, at a given SNR we first simulate each equivalent subchannel to estimate the marginal output conditional densities observed at each level/stage of decoding. For each of these densities, we then optimize an LDPC degree sequence for maximal rate under the constraint that iterative decoding converges. The results of these optimizations (for a maximum left-degree of 30) are shown in Fig. 5, and an example degree sequence at rate 0.9 is presented in Table II. For most SNRs, the iterative-decoding threshold is within 0.2 dB of $I_{\text{i.u.d.}}$. In Fig. 6, we plot simulation results for codes constructed from the degree sequences in Table II, subject to the constraint of no length-4 cycles in the bipartite graph representation. A 2-level system with component block length $N = 10^6$ achieves a BER $= 10^{-5}$ within 0.1 dB of the iterative-decoding threshold, and within 0.3 dB of $I_{\text{i.u.d.}}$. Naturally, one might expect to reduce this gap by going to m-level systems with larger values of m, and presumably as in [51] by considering code ensembles with larger maximum left degrees.

Note that for this particular case of a two-level system on the dicode channel, the design of the second code is simplified slightly since the second equivalent subchannel is a binary-input Gaussian channel with $E_s/N_0 = 1/(2\sigma^2)$ (Corollary 1). This is also apparent in Fig. 4(b), where $R_2^{(2)}$ equals the capacity of a binary-input AWGN channel.

Example 2: EPR4 channel, $h(D) = (1 + D - D^2 - D^3)/2$.

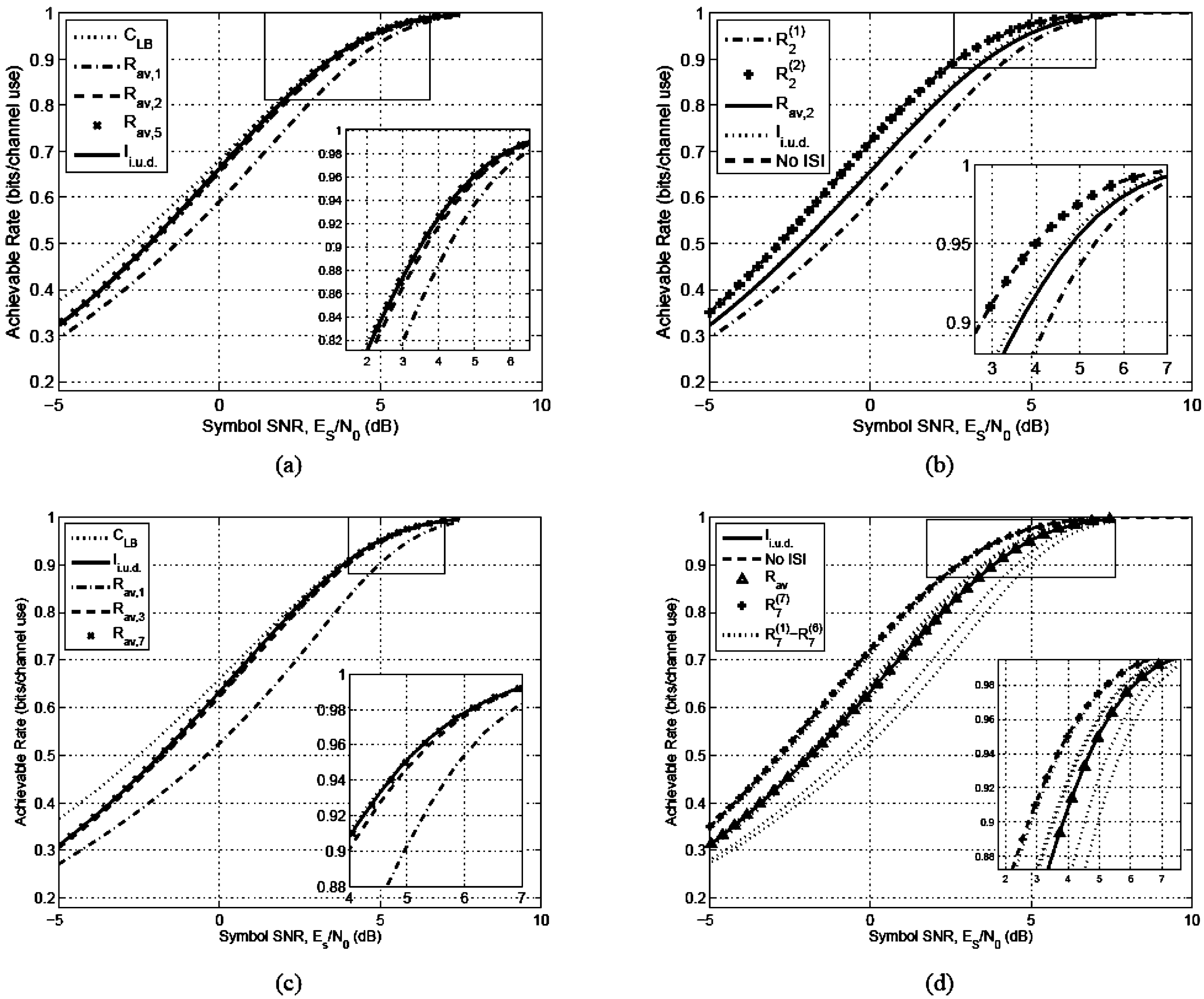

Fig. 4. (a) Achievable rates for the overall MLC/MSD system on the dicode channel with $m = 1, 2$, and 5 levels. (b) Achievable rates of component codes for a 2-level system on the dicode channel. Notice that, in accordance with Corollary 1, $R_2^{(2)}$ equals the binary-input AWGN channel capacity. (c) Achievable rates for the overall MLC/MSD system on the EPR4 channel with $m = 1, 3$, and 7 levels. The curves for $R_{\mathrm{av},7}$ and $I_{\mathrm{i.u.d.}}$ are almost identical in this plot. (d) Achievable rates of component codes for a 7-level system on the EPR4 channel.

TABLE I
OVERALL ACHIEVABLE MLC/MSD SYSTEM RATES ON THE DICODE AND EPR4 CHANNELS, FOR VARIOUS m

Channel	σ	$R_{\mathrm{av},1}$	$R_{\mathrm{av},2}$	$R_{\mathrm{av},3}$	$R_{\mathrm{av},4}$	$R_{\mathrm{av},5}$	$R_{\mathrm{av},6}$	$R_{\mathrm{av},7}$	$I_{\mathrm{i.u.d.}}$
dicode	0.4742	0.854	0.892	0.900	–	–	–	–	0.9
	0.5722	0.732	0.792	0.798	0.8	–	–	–	0.8
	0.6691	0.626	0.694	0.699	0.7	–	–	–	0.7
EPR4	0.4546	0.819	0.847	0.892	0.893	0.898	0.898	0.899	0.9
	0.5460	0.688	0.725	0.789	0.789	0.797	0.798	0.799	0.8
	0.6387	0.584	0.621	0.687	0.695	0.695	0.699	0.699	0.7

In Fig. 4(c), we show achievable rates for various m-level systems on the EPR4 channel. In this case, we find that using 3-levels is sufficient for closely approaching $I_{\mathrm{i.u.d.}}$ (also apparent in Table I). This is especially interesting since m is shorter than the channel impulse response length. For the achievable rates of a 7-level system in Fig. 4(d), the component rate $R_7^{(7)}$ equals the capacity of a binary-input AWGN channel, and $R_7^{(i)} > I_{\mathrm{i.u.d.}}$ for $i > 3$. These observations are both consistent with the analysis in Section III. We also notice that $R_7^{(3)}, \ldots, R_7^{(5)}$ are all very close to $I_{\mathrm{i.u.d.}}$.

We then choose $m = 3$ and proceed to design component LDPC codes. Fig. 7 shows the iterative-decoding thresholds for various 3-level systems with optimized LDPC code ensembles, and a rate–0.9 degree sequence is given in Table II. The corresponding simulation results for codes chosen from this ensemble are shown in Fig. 6, where we find that with a component

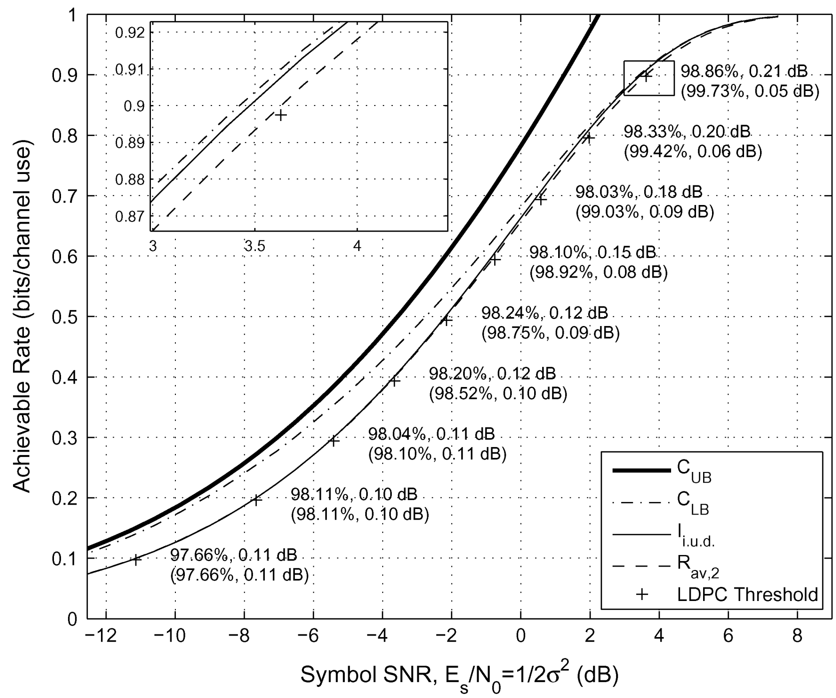

Fig. 5. Iterative-decoding thresholds for 2-level MLC/MSD systems with component LDPC codes on the dicode channel. The maximum left degree for the LDPC codes is 30. For each point, the percentage within $I_{\rm i.u.d.}$ and the decibel gap from the $I_{\rm i.u.d.}$ threshold are shown, and in parenthesis are the percentage within $R_{\rm av,2}$ and the decibel gap from the $R_{\rm av,2}$ threshold. The capacity lower bound is the mutual information rate for an optimized 8-state binary Markov input process [2], [3], and the upper bound corresponds to the Gaussian-input water-filling capacity [9, p. 256].

TABLE II
OPTIMIZED 2-LEVEL SYSTEM FOR THE DICODE CHANNEL AND 3-LEVEL SYSTEM FOR THE EPR4 CHANNEL

2-levels on Dicode Channel $R_{\rm av,2}=0.8975$				3-levels on EPR4 Channel $R_{\rm av,3}=0.8974$					
1st level		2nd level		1st level		2nd level		3rd level	
$R(\mathcal{C}^{(1)})=0.8583$		$R(\mathcal{C}^{(2)})=0.9366$		$R(\mathcal{C}^{(1)})=0.8258$		$R(\mathcal{C}^{(2)})=0.9224$		$R(\mathcal{C}^{(3)})=0.9439$	
i	$\lambda_i^{(1)}$		$\lambda_i^{(2)}$		$\lambda_i^{(1)}$		$\lambda_i^{(2)}$		$\lambda_i^{(3)}$
2	0.1392		0.1195		0.1395		0.1195		0.1195
3	0.2127		0.2265		0.1949		0.2160		0.2327
6					0.0167				
7	0.0930		0.1390		0.2197		0.0962		0.0594
8	0.1792		0.0802				0.1867		0.2426
9	0.0263								
10			0.1007						
19			0.0107						0.0535
29									0.2423
30	0.3495		0.3233		0.4292		0.3816		0.0500
$\rho_{36}^{(1)}$	0.0081	$\rho_{84}^{(2)}$	0.4394	$\rho_{31}^{(3)}$	0.6583	$\rho_{70}^{(3)}$	0.0263	$\rho_{94}^{(3)}$	0.5926
$\rho_{37}^{(1)}$	0.9919	$\rho_{85}^{(2)}$	0.5606	$\rho_{32}^{(3)}$	0.3417	$\rho_{71}^{(3)}$	0.9737	$\rho_{95}^{(3)}$	0.4074

block length of $N=10^6$, a BER $=10^{-5}$ is achieved within 0.1 dB of the iterative-decoding threshold, and within 0.2 dB of $I_{\rm i.u.d.}$.

VI. CONCLUSION

By focusing on the marginal conditional output densities at each stage of MSD, we are able to determine achievable rates for binary ISI channels. We then design and construct LDPC codes which allow MLC/MSD systems to approach these rates. The derivation of the achievable rates, along with a Monte Carlo method for computing them, is applicable to general stationary ergodic channels. In the case of binary ISI channels, we further show that these achievable rates for MLC/MSD converge to the i.u.d. information rate as the number of coding levels and decoding stages increases. For the design of LDPC codes, we show that the application of [30], [8] to the marginal conditional output densities can be used to design LDPC codes which allow MLC/MSD to approach the aforementioned achievable rates. This follows from a straightforward extension of the concentration theorem in [30], [8].

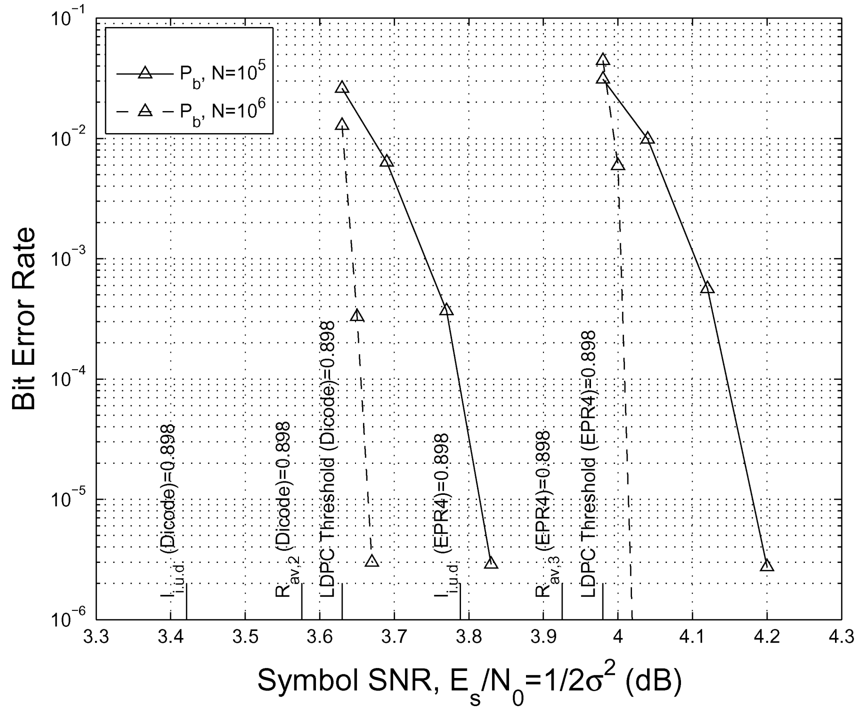

Fig. 6. Performance of optimized rate–0.9 MLC/MSD systems: (left) 2-level system for the dicode channel; (right) 3-level system for the EPR4 channel. The degree sequences for the codes are shown in Table II. Simulation reflects the BER of the all-zero codeword averaged over randomly chosen coset vectors and noise realizations.

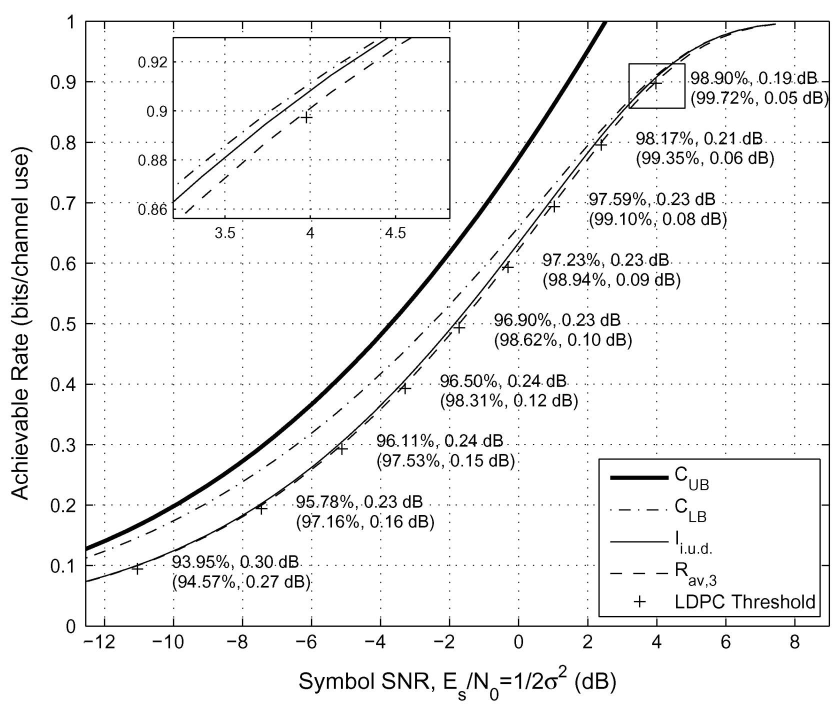

Fig. 7. Iterative-decoding thresholds for 3-level MLC/MSD systems with component LDPC codes on the EPR4 channel. The maximum left degree for the LDPC codes is 30. For each point, the percentage within $I_{\rm i.u.d.}$ and the decibel gap from the $I_{\rm i.u.d.}$ threshold are shown, and in parenthesis are the percentage within $R_{\rm av,3}$ and the decibel gap from the $R_{\rm av,3}$ threshold. The capacity lower bound is the mutual information rate for an optimized 8-state binary Markov input process [2], [3], and the upper bound corresponds to the Gaussian-input water-filling capacity [9, p. 256].

For the dicode and EPR4 channels, which arise in the context of magnetic recording, our achievable rate analysis shows that MLC/MSD operation within 0.2 dB of $I_{\rm i.u.d.}$ can be obtained with as few as two and three levels, respectively. Furthermore, we demonstrate a 2-level system on the dicode channel with optimized component LDPC codes for which the iterative-decoding threshold is within 0.2 dB of $I_{\text{i.u.d.}}$, and for which simulations at a component code block length of $N = 10^6$ achieve a BER $= 10^{-5}$ within 0.3 dB of $I_{\text{i.u.d.}}$. For a similarly optimized 3-level system on the EPR4 channel, we observe an iterative-decoding threshold within 0.1 dB of $I_{\text{i.u.d.}}$, and a BER $= 10^{-5}$ within 0.2 dB of $I_{\text{i.u.d.}}$.

We note that, although not all of the analysis here was extended to general finite-state or stationary ergodic channels, these methods have still been found to be effective in cases other than the binary ISI channel. Such examples include MLC/MSD systems on two-dimensional binary ISI channels [52], which lead to lower bounds and thresholds very close to the i.u.d. information rate (e.g., [53]), as well as concatenated coding with inner trellis codes and an outer MLC/MSD system, which can be used to approach the channel capacity of binary ISI channels [54], [55]. (Details can be found in [45].) Generalizations for channels where the channel state is not necessarily determined by the inputs can be found in [37], [38].

Appendix I
Coding Theorem With Mismatched Decoders

When proving Theorem 1 in Section III-A, we mentioned that one could prove achievability once the limit in (8) is established. We now give details to confirm this. Incidentally, this achievability result was also proven by Ganti *et al.* [56], but here we provide an alternative proof both for completeness and also because interestingly our approach demonstrates how the typical-set decoding analysis of [9] is powerful enough to be extended to mismatched decoders.

Theorem 5: Let $\{X_t\}$ and $\{Y_t\}$ be the input and output processes, respectively, of a channel with a finite input alphabet $\mathbb{X}$ and continuous output alphabet $\mathbb{Y}$. Let $\{q_N(\boldsymbol{x}^N|\boldsymbol{y}^N), N \geq 1\}$ be a sequence of well-defined conditional probability functions on these inputs and outputs, i.e., for each $N \geq 1$, i) $0 \leq q_N(\boldsymbol{x}^N|\boldsymbol{y}^N) \leq 1$ for all possible $\boldsymbol{x}^N$ and $\boldsymbol{y}^N$, and ii) $\sum_{\boldsymbol{x}^N} q_N(\boldsymbol{x}^N|\boldsymbol{y}^N) = 1$ for all possible $\boldsymbol{y}^N$. Lastly, assume that

$$\lim_{N\to\infty} \frac{1}{N} \log_2 \frac{q_N(\boldsymbol{X}^N|\boldsymbol{Y}^N)}{p(\boldsymbol{X}^N)} = R' \quad \text{(a.s.)} \tag{14}$$

where $p(\boldsymbol{x}^N) = P(\boldsymbol{X}^N = \boldsymbol{x}^N)$. (We call R' the *mismatch rate*.)

Then, any rate $R < R'$ can be achieved with decoders based only on the conditional densities, $\{q_N\}$.

Proof: First, for any $\delta > 0$, let us define the set

$$A_\delta^N = \Bigg\{ \boldsymbol{x}^N \in \mathbb{X}^N, \\ \boldsymbol{y}^N \in \mathbb{R}^N : \left| \frac{1}{N} \log_2 \frac{q(\boldsymbol{x}^N|\boldsymbol{y}^N)}{p(\boldsymbol{x}^N)} - R' \right| < \delta \Bigg\}. \tag{15}$$

Using the concept of typical-set decoding [9], we can define a decoder to take the received sequence $\boldsymbol{y}^N$ and return all code vectors $\boldsymbol{x}^N$ for which the pair $(\boldsymbol{x}^N, \boldsymbol{y}^N) \in A_\delta^N$. Then, the achievability of R' using this decoder is possible if it can be shown that, as $N \to \infty$, the probability of an atypical (or undecodable) event satisfies $\alpha_N = P((\boldsymbol{X}^N, \boldsymbol{Y}^N) \notin A_\delta^N) \to 0$, and also the probability of a miscorrection satisfies

$$\beta_N = P\left((\tilde{\boldsymbol{X}}^N, \boldsymbol{Y}^N) \in A_\delta^N \right) \leq 2^{-N(R'-\delta)}$$

where $\{\tilde{X}_t\}$ is defined to have the same distribution as $\{X_t\}$ but is independent of both $\{X_t\}$ and $\{Y_t\}$. This is because the average probability of decoding error over the Shannon ensemble with code rate R equals $\alpha_N + \beta_N 2^{NR}$, which goes to zero whenever $R < R' - \delta$.

The first condition, $P((\boldsymbol{X}^N, \boldsymbol{Y}^N) \notin A_\delta^N) \to 0$, follows immediately from the assumption in (14). To show that the second property holds, we define the subset $A_\delta^N(\boldsymbol{x}^N)$ for each $\boldsymbol{x}^N \in \mathbb{X}^N$ as

$$A_\delta^N(\boldsymbol{x}^N) = \left\{ \boldsymbol{y}^N : (\boldsymbol{x}^N, \boldsymbol{y}^N) \in A_\delta^N \right\}.$$

We can then deduce that

$$\begin{aligned} P&\left((\tilde{\boldsymbol{X}}^N, \boldsymbol{Y}^N) \in A_\delta^N \right) \\ &= \sum_{\boldsymbol{x}^N \in \mathbb{X}^N} P(\tilde{\boldsymbol{X}} = \boldsymbol{x}^N) P\left(\boldsymbol{Y}^N \in A_\delta^N(\boldsymbol{x}^N) | \tilde{\boldsymbol{X}} = \boldsymbol{x}^N \right), \\ &= \sum_{\boldsymbol{x}^N \in \mathbb{X}^N} P\left(\boldsymbol{X} = \boldsymbol{x}^N \right) P\left(\boldsymbol{Y}^N \in A_\delta^N(\boldsymbol{x}^N) \right), \end{aligned} \tag{16}$$

because, by assumption, $\boldsymbol{Y}^N$ is independent of $\tilde{\boldsymbol{X}}^N$, and $P(\tilde{\boldsymbol{X}} = \boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x})$. From (15) we have

$$p(\boldsymbol{x}^N) \leq q(\boldsymbol{x}^N|\boldsymbol{y}^N) 2^{-NR'+N\delta}$$

for all $\boldsymbol{y}^N \in A_\delta^N(\boldsymbol{x}^N)$. Therefore, we can bound

$$\begin{aligned} p(\boldsymbol{x}^N)&P\left(\boldsymbol{Y}^N \in A_\delta^N(\boldsymbol{x}^N) \right) \\ &= \int_{\boldsymbol{y}^N \in A_\delta^N(\boldsymbol{x}^N)} p(\boldsymbol{x}^N) f(\boldsymbol{y}^N) dy_1 \ldots dy_N \\ &\leq 2^{-NR'+N\delta} \int_{\boldsymbol{y}^N \in A_\delta^N(\boldsymbol{x}^N)} q(\boldsymbol{x}^N|\boldsymbol{y}^N) f(\boldsymbol{y}^N) dy_1 \ldots dy_N \\ &\leq 2^{-NR'+N\delta} \int_{\boldsymbol{y}^N \in \mathbb{R}^N} q(\boldsymbol{x}^N|\boldsymbol{y}^N) f(\boldsymbol{y}^N) dy_1 \ldots dy_N \end{aligned}$$

(assuming the density $f(\boldsymbol{y}^N)$ exists). We can substitute the preceding bound into (16), yielding

$$\begin{aligned} P&\left((\tilde{\boldsymbol{X}}^N, \boldsymbol{Y}^N) \in A_\delta^N \right) \\ &\leq \sum_{\boldsymbol{x}^N \in \mathbb{X}^N} 2^{-NR'+N\delta} \int_{\boldsymbol{y}^N \in \mathbb{R}^N} q(\boldsymbol{x}^N|\boldsymbol{y}^N) f(\boldsymbol{y}^N) dy_1 \ldots dy_N \end{aligned}$$

and then exchange the order of summation and integration because the sum is finite, resulting in

$$2^{-NR'+N\delta} \int_{\boldsymbol{y}^N \in \mathbb{R}^N} f(\boldsymbol{y}^N) dy_1 \ldots dy_N = 2^{-NR'+N\delta}. \quad \square$$

To apply this theorem specifically to the ith equivalent subchannel of Theorem 1, set

$$q_N(\boldsymbol{x}^N|\boldsymbol{y}^N) = \prod_{j=1}^{N} P\left(C_j^{(i)} = x_j | L_j^{(i)} = y_j \right).$$

Traditionally, for the case of mismatched decoders like these where the decoder uses only the marginal channel density, achievability is usually proved by interleaving many codes so that each code effectively sees a memoryless channel. Here, the theorem allows us to include a larger class of mismatched decoders.

APPENDIX II
PROOF OF COROLLARY 1

Proof: Recall from the multistage decoding structure that the mth stage has decisions for all previous $(m-1)$ interleaves. So, if j is any interior point on the mth equivalent subchannel, then from Definition 1 we know that the APP detector has perfect information for all $(m-1)$ previous and all $(m-1)$ subsequent channel inputs when detecting X_{jm}. More importantly, if $m > \nu$, we can use this information to completely subtract the ISI from $\boldsymbol{Y}_k^{k+\nu}$, leaving $(\nu+1)$ independent observations of X_k corrupted by AWGN terms; i.e.,

$$Y'_{k+i} = h_i X_k + N_{k+i}, \qquad \text{for } i = 0, \ldots, \nu.$$

The classic optimal detector for this problem is the maximal ratio combiner [57, p. 779], which forms the sufficient statistic

$$Z_k = \sum_{i=0}^{\nu} h_i Y'_{k+i} = ||\boldsymbol{h}||^2 X_k + ||\boldsymbol{h}|| N'_k$$

where $N'_k \sim \mathcal{N}(0,\sigma^2)$, and $||\boldsymbol{h}||^2 = \sum_{i=0}^{\nu} h_i^2$. Notice that this describes a binary-input AWGN channel with SNR $||\boldsymbol{h}||^2/\sigma^2$. In fact, the mth equivalent subchannel reduces to this channel because

$$\begin{aligned} L_j^{(m)} &= \log \frac{P\left(Y'_k \ldots Y'_{k+\nu} | X_k = 1\right)}{P\left(Y'_k \ldots Y'_{k+\nu} | X_k = -1\right)} \\ &= \frac{1}{2\sigma^2} \sum_{i=0}^{\nu} -\left(Y'_{k+i} - h_i\right)^2 + \left(Y'_{k+i} + h_i\right)^2 \\ &= \frac{2}{\sigma^2} \sum_{i=0}^{\nu} h_i Y'_{k+i} = \frac{2}{\sigma^2} Z_k. \end{aligned}$$

Here we see that $L_j^{(m)}$ corresponds to the output of a *normalized* binary-input Gaussian channel with the same capacity [48]. □

ACKNOWLEDGMENT

The authors wish to thank the referees for their comments and suggestions which have greatly improved the presentation of this paper. The authors also wish to thank the Associate Editor for his help during the review process.

REFERENCES

[1] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260–2299, Oct. 1998.
[2] D. Arnold and H. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. IEEE Int. Conf. Communications*, Helsinki, Finland, Jun. 2001, pp. 2692–2695.
[3] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. IEEE Global Telecommunications Conf.*, San Antonio, TX, Nov. 2001, pp. 2992–2996.
[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. Information Theory*, Washington, DC, Jun. 2001, p. 283.
[5] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
[6] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo equalization," *Europ. Trans. Telecommun. Related Technol.*, vol. 6, pp. 507–511, Sept.–Oct. 1995.
[7] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.
[8] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[10] A. R. Barron, "The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem," *Ann. Probab.*, vol. 13, pp. 1292–1303, Nov. 1985.
[11] W. Hirt, "Capacity and information rates of discrete-time channels with memory," Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1988.
[12] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov. 1991.
[13] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.
[14] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *Proc. IEEE Global Telecommunications Conf.*, San Antonio, TX, Nov. 2001, pp. 2997–3001.
[15] P. O. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proc. IEEE Information Theory Workshop*, Cairns, Australia, Sep. 2001, pp. 147–149.
[16] S. Yang, A. Kavčić, and S. Tatikonda, "Delayed feedback capacity of finite-state machine channels: Upper bounds on the feedforward capacity," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 290.
[17] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[18] W. E. Ryan, L. L. McPheters, and S. W. McLaughlin, "Combined turbo coding and turbo equalization for PR4-equalized Lorentzian channels," in *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 1998, pp. 489–493.
[19] T. Souvignier, M. Öberg, P. H. Siegel, R. E. Swanson, and J. K. Wolf, "Turbo decoding for partial response channels," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1297–1308, Aug. 2000.
[20] M. Öberg and P. H. Siegel, "Performance analysis of turbo-equalized dicode partial-response channel," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 436–444, Mar. 2001.
[21] J. L. Fan, A. Friedmann, E. Kurtas, and S. W. McLaughlin, "Low density parity check codes for magnetic recording," in *Proc. 37th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 1999, pp. 1314–1323.
[22] M. Tüchler, C. Weiß, E. Eleftheriou, A. Dholakia, and J. Hagenauer, "Application of high-rate tail-biting codes to generalized partial response channels," in *Proc. IEEE Global Telecommunications Conf.*, San Antonio, TX, Nov. 2001, vol. 5, pp. 2966–2971.
[23] K. Narayanan, "Effect of precoding on the convergence of turbo equalization for partial response channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 686–698, Apr. 2001.
[24] D. Doan and K. R. Narayanan, "Some new results on the design of codes for inter-symbol interference channels based on convergence of turbo equalization," in *Proc. IEEE Int. Conf. Communications*, New York, Apr. 2002, pp. 1873–1877.
[25] A. Thangaraj and S. W. McLaughlin, "Thresholds and scheduling for LDPC-coded partial-response channels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2307–2309, Sep. 2002.

[26] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," in *AEÜ Int. J. Electron. and Commun.*, Nov. 2000, vol. 54, pp. 389–398.
[27] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1636–1652, Jul. 2003.
[28] N. Varnica and A. Kavčić, "Optimized LDPC codes for partial response channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun. 2002, p. 197.
[29] ——, "Optimized low-density parity-check codes for partial response channels," *IEEE Commun. Lett.*, vol. 7, no. 4, pp. 168–170, Apr. 2003.
[30] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
[31] B. F. Filho, M. A. Herro, and D. J. Costello, Jr., "A multilevel approach to constructing trellis-matched codes for binary-input partial-response channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2582–2591, Nov. 1999.
[32] A. V. Kuznetsov, "Coded modulation schemes with turbo and list detection for high-order partial response channels," *IEEE Trans. Magn.*, vol. 37, no. 2, pp. 729–736, Mar. 2001.
[33] J. A. Miller and J. K. Wolf, "High code rate ECC design for partial response systems," *IEEE Trans. Magn.*, vol. 37, no. 2, pp. 704–707, Mar. 2001.
[34] X. Ma, N. Varnica, and A. Kavčić, "Matched information rate codes for binary ISI channels," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 269.
[35] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "On approaching the capacity of finite-state intersymbol interference channels," in *Information, Coding, and Mathematics: Proc. Workshop Honoring Prof. Bob McEliece on His 60th Birthday*, M. Blaum, P. G. Farrell, and H. C. A. van Tilborg, Eds. Boston, MA: Kluwer Academic, 2002, pp. 365–378.
[36] K. R. Narayanan and N. Nangare, "A BCJR-DFE based receiver for achieving near capacity performance on ISI channels," in *Proc. 42nd Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004, pp. 763–772.
[37] T. Li and O. M. Collins, "A successive decoding strategy for channels with memory," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 239–243.
[38] ——, "A successive decoding strategy for channel with memory," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 628–646, Feb. 2007.
[39] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.
[40] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, Sep. 2003.
[41] B. M. Kurkoski, P. H. Siegel, and J. K. Wolf, "Joint message-passing decoding of LDPC codes and partial-response channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1410–1422, Jun. 2002.
[42] F. Le Gland and L. Mevel, "Exponential forgetting and geometric ergodicity in hidden Markov models," *Math. Control Signals Syst.*, vol. 13, pp. 63–93, Jul. 2000.
[43] ——, "Basic properties of the projective product with application to products of column-allowable nonnegative matrices," *Math. Control Signals Syst.*, vol. 13, pp. 41–62, Jul. 2000.
[44] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.
[45] J. B. Soriaga, "On near-capacity code design for partial-response channels," Ph.D. dissertation, Univ. Calif., San Diego, La Jolla, Mar. 2005.
[46] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York: Springer-Verlag, 1988.
[47] A. Kavčić, X. Ma, and N. Varnica, "Matched information rate codes for partial response channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 973–989, Mar. 2005.
[48] S.-Y. Chung, "On the construction of some capacity-approaching coding schemes," Ph.D. dissertation, MIT, Cambridge, MA, 2001.
[49] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Multilevel coding with low-density parity-check component codes," in *Proc. IEEE Global Telecommunications Conf.*, San Antonio, TX, Nov. 2001, pp. 1016–1020.
[50] R. L. Urbanke and T. J. Richardson, Modern Coding Theory, 2003 [Online]. Available: http://lthcwww.epfl.ch/mct, Cambridge Univ. Press, preprint.
[51] S. Chung, G. D. Forney, Jr., T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.
[52] J. B. Soriaga, M. Marrow, P. H. Siegel, and J. K. Wolf, "On achievable rates of multistage decoding on two-dimensional ISI channels," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sept. 2005, pp. 1348–1352.
[53] J. Chen and P. Siegel, "On the symmetric information rate of two-dimensional finite state ISI channels," in *Proc. IEEE Information Theory Workshop*, Paris, France, Mar. 2003, pp. 320–323.
[54] J. B. Soriaga and P. H. Siegel, "Near-capacity coding systems for partial-response channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, June 2004, p. 267.
[55] J. B. Soriaga and P. H. Siegel, "On distribution shaping codes for partial-response channels," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 468–477.
[56] A. Ganti, A. Lapidoth, and İ. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.
[57] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.