Bit-Interleaved Coded Modulation

Giuseppe Caire, *Member, IEEE*, Giorgio Taricco, *Member, IEEE*, and Ezio Biglieri, *Fellow, IEEE*

***Abstract*—It has been recently recognized by Zehavi that the performance of coded modulation over a Rayleigh fading channel can be improved by bit-wise interleaving at the encoder output, and by using an appropriate soft-decision metric as an input to a Viterbi decoder. The goal of this paper is to present in a comprehensive fashion the theory underlying bit-interleaved coded modulation, to provide tools for evaluating its performance, and to give guidelines for its design.**

***Index Terms*—Bit-interleaving, channel capacity, coded modulation, cutoff rate, fading channel.**

I. Introduction and Motivations

EVER since 1982, when Ungerboeck published his landmark paper on trellis-coded modulation [19], it has been generally accepted that modulation and coding should be combined in a single entity for improved performance. Of late, the increasing interest for mobile-radio channels has led to the consideration of coded modulation for fading channels. Thus at first blush it seemed quite natural to apply the same "Ungerboeck's paradigm" of keeping coding combined with modulation even in a situation (the Rayleigh fading channel) where the code performance depends strongly, rather than on the minimum Euclidean distance of the code, on its minimum Hamming distance (the "code diversity"). Several results followed this line of thought, as documented by a considerable body of work aptly summarized and referenced in [14] (see also [5, Ch. 10]). Under the assumption that the symbols were interleaved with a depth exceeding the coherence time of the fading process, new codes were designed for the fading channel so as to maximize their diversity. This implied in particular that parallel transitions should be avoided in the code, and that any increase in diversity would be obtained by increasing the constraint length of the code.

A notable departure from Ungerboeck's paradigm was the core of [24]. Schemes were designed aimed at keeping as their basic engine an off-the-shelf Viterbi decoder for the *de facto* standard, 64-state rate-$1/2$ convolutional code. This implied giving up the joint decoder/demodulator in favor of two separate entities.

Based on the latter concept, Zehavi [26] recognized that the code diversity, and hence the reliability of coded modulation over a Rayleigh fading channel, could be further improved. Zehavi's idea was to make the code diversity equal to the smallest number of distinct *bits* (rather than *channel symbols*) along any error event. This is achieved by bit-wise interleaving at the encoder output, and by using an appropriate soft-decision bit metric as an input to the Viterbi decoder. Further results along this line were recently reported in [2], [13], and [1] (for different approaches to the problem of designing coded-modulation schemes for the fading channels see [20] and [6]).

This paper is based on Zehavi's findings, and in particular on his result, rather surprising *a priori*, that on some channels there is a downside to combining demodulation and decoding. Our goal is to present in a comprehensive fashion the theory underlying bit-interleaved coded modulation (BICM) and to provide a general information-theoretical framework for this concept. This analysis also yields tools for evaluating the performance of BICM (with bounds to error probabilities tighter than those previously known) as well as guidelines for its design.

Definitions and channel model are first introduced (Section II). Next, the information-theoretical foundations of BICM are laid in Section III by evaluating in general the capacity and the cutoff rate of bit-interleaved channels. Section IV is devoted to error analysis: various approximations and bounds are introduced, and the effect of the choice of signal labeling is discussed. In particular, the asymptotic optimality of Gray labeling in conjunction with BICM is showed. Design criteria are pointed out in Section V, where a number of examples are also shown. Conclusions are summarized in Section VI, where the main themes of this paper are reprised.

II. System Model

In this section we recall the baseline model of coded modulation (CM) and introduce the model of BICM.

Before proceeding further, let us stipulate a terminological convention: hereafter we shall use the term "bit" to denote a binary digit, and "information bit" to denote the binary information unit. Thus for example, a string of m bits is a sequence of symbols 0's and 1's. A random vector with m uniform independent and identically distributed (i.i.d.) components has an entropy of m information bits. Moreover, lower case symbols denote scalar quantities, boldface symbols denote vectors, and underlined boldface symbols (e.g., $\underline{\boldsymbol{c}}$) denote sequences of scalars or of vectors.

The CM and BICM models are represented by the block diagram of Fig. 1. The building blocks of both schemes are 1) an encoder (ENC); 2) an interleaver π; 3) a modulator, modeled by a labeling map μ and a signal set $\mathcal{X}$, i.e., a finite set of points in the complex N-dimensional Euclidean space $\mathbb{C}^N$; 4) a stationary finite-memory vector channel whose transition probability density function $p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ may depend on a vector parameter $\boldsymbol{\theta}$; 5) a demodulator (DEM), which in the present scenario plays the role of a branch metric computer; 6)

Manuscript received August 11, 1996; revised June 1, 1997. This work was supported by the Italian Space Agency (ASI).

The authors are with Politecnico di Torino, I-10129 Torino, Italy.

Publisher Item Identifier S 0018-9448(98)02360-8.

0018–9448/98$10.00 © 1998 IEEE

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

LG Electronics, Inc. v. Constellation Designs, LLC IPR2023-00319 Constellation Exhibit 2015
Page 1 of 20

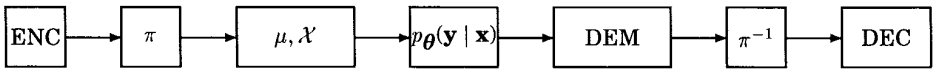

Fig. 1. Block diagram of transmission with coded modulation (CM) and bit-interleaved coded modulation (BICM). In the case of CM, π denotes interleaving at the symbol level. In the case of BICM, π denotes interleaving at the bit level.

a branch metric deinterleaver π^{-1}; 7) a decoder (DEC). The balance of this section is devoted to a detailed description of these blocks for CM and BICM.

A. Vector Channel Model

We consider a vector channel characterized by a family of transition probability density functions (pdf)

$$\{p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) \ : \ \boldsymbol{\theta} \in \mathbb{C}^{N'}; \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N\} \tag{1}$$

parameterized by a complex-valued vector $\boldsymbol{\theta}$ representing the channel state. We assume that $\boldsymbol{\theta}$ is independent of the channel input $\boldsymbol{x}$, and that, conditionally on the sequence $\underline{\boldsymbol{\theta}}$, the channel is memoryless, i.e.,

$$p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}) = \prod_k p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid \boldsymbol{x}_k). \tag{2}$$

Moreover, $\underline{\boldsymbol{\theta}}$ is assumed to be a stationary, finite memory random process. More precisely, we assume that there exists an integer $\nu \geq 0$ such that, for all r-tuples $\nu < k_1 < \cdots < k_r$ and for all n-tuples $j_1 < \cdots < j_n \leq 0$, the sequences $(\boldsymbol{\theta}_{k_1}, \cdots, \boldsymbol{\theta}_{k_r})$ and $(\boldsymbol{\theta}_{j_1}, \cdots, \boldsymbol{\theta}_{j_n})$ are statistically independent.

By this model we can represent a large number of typical communication channels such as the additive white Gaussian noise (AWGN) channel ($\boldsymbol{\theta}$ = constant), the AWGN channel with random phase ($\boldsymbol{\theta}$ is the residual phase due to imperfect carrier phase recovery) and frequency nonselective slow-fading channels [17] ($\boldsymbol{\theta}$ describes the multiplicative fading process). However, the model cannot take into account intersymbol interference (ISI), or frequency selectivity in fading channels since, in these cases, the channel state depends on the input sequence.

B. Coded Modulation

A CM scheme can be viewed as the concatenation of an encoder ENC for a code $\mathcal{C}$ defined over a label alphabet $\mathcal{A}$ with an N-dimensional memoryless modulator over a signal set $\mathcal{X} \subseteq \mathbb{C}^N$ [18], [19]. We assume that $\mathcal{A}$ and $\mathcal{X}$ are finite and $|\mathcal{A}| = |\mathcal{X}| = M$. The transmitted code sequence $\underline{\boldsymbol{c}} \in \mathcal{C}$ is mapped onto the signal sequence $\underline{\boldsymbol{x}}$ by the (one-to-one) labeling map $\mu : \mathcal{A} \to \mathcal{X}$, acting componentwise.

If the code $\mathcal{C}$ is designed to correct random errors, it may be convenient to introduce a *symbol interleaver* π between the encoder ENC and the mapper μ. For most of this paper we shall assume that π is an ideal interleaver (i.e., with infinite depth and completely random). Thus the interleaving/deinterleaving blocks can be excluded from consideration provided that the sequence $\underline{\boldsymbol{\theta}}$ is i.i.d., and hence characterized by its first-order pdf.

Let $\underline{\boldsymbol{y}} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_k, \cdots)$ denote the channel output sequence resulting from the transmission of $\underline{\boldsymbol{x}}$. Optimum decoding depends on the channel state information (CSI) available at the receiver. Here we consider the two extreme cases of *perfect* CSI and no CSI.

1) Detection with Perfect CSI: By assuming that the receiver has perfect knowledge of the values taken on by each $\boldsymbol{\theta}_k$, the set of maximum-likelihood (ML) symbol metrics is given by

$$\{\log p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid \boldsymbol{z})\}_{\boldsymbol{z} \in \mathcal{X}}. \tag{3}$$

After metric deinterleaving (denoted by π^{-1} in Fig. 1), ML decision about the transmitted code sequence is made according to the rule

$$\hat{\underline{\boldsymbol{c}}} = \arg\max_{\underline{\boldsymbol{c}} \in \mathcal{C}} \sum_k \log p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid \mu(c_k)) \tag{4}$$

where the summation runs over the whole sequence. Note that (4) is the ML decision rule also with finite-depth interleaving, or with no interleaving at all, since conditionally on $\underline{\boldsymbol{\theta}}$, the channel is memoryless (see (2)).

2) Detection with No CSI: The optimum receiver structure in the absence of CSI is different. A major difference is that, $\underline{\boldsymbol{\theta}}$ being unknown, the channel is not memoryless.[1] However, with ideal interleaving, for all finite-index sets $\mathcal{K}$

$$p(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}) = E_{\underline{\boldsymbol{\theta}}}[p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}})] = \prod_{k \in \mathcal{K}} p(\boldsymbol{y}_k \mid \boldsymbol{x}_k)$$

where we defined a new average transition pdf as

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = E_{\boldsymbol{\theta}}[p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})]. \tag{5}$$

Then, in the case of no CSI and ideal interleaving, the ML symbol metric and decision rule are obtained by using $p(\boldsymbol{y} \mid \boldsymbol{x})$ instead of $p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})$ in (3) and in (4), respectively.

Example 1: Ungerboeck Trellis-Coded Modulation (TCM): The first and best known construction of CM [19] is obtained by choosing for $\mathcal{A}$ the set $\{0,1\}^m$, for $\mathcal{C}$ a binary convolutional code of rate k_c/m, and for μ the *set partitioning* labeling, i.e., a binary labeling obtained by an m-level binary tree partition of $\mathcal{X}$ with nondecreasing minimum Euclidean distance of the signal subsets at subsequent partition levels [19], [5]. An example of set partitioning labeling for 16QAM is given in Fig. 2(a). We define the *spectral efficiency* of this construction to be $R = k_c/N$ information bits per complex dimension (bit/dim).

C. Bit-Interleaved Coded Modulation

In general, BICM can be obtained by concatenating an encoder ENC for a binary code $\mathcal{C}$, with an N-dimensional memoryless modulator over a signal set $\mathcal{X} \subseteq \mathbb{C}^N$ of size $|\mathcal{X}| = M = 2^m$, through a bit interleaver π and a one-to-one

[1] Noncoherent detection is a typical case: if the carrier phase is a slowly varying process, performance enhancement is achievable, even with no coding, by detecting blocks of signals [12].

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

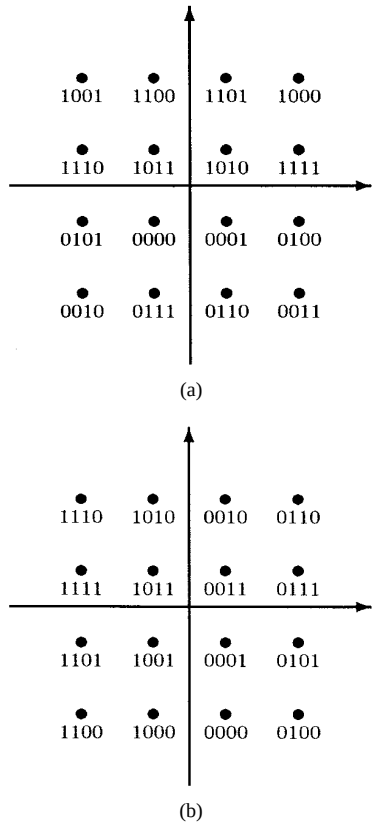

Fig. 2. 16QAM signal set with (a) a suitable set partitioning labeling and (b) Gray labeling.

binary labeling map $\mu : \{0,1\}^m \to \mathcal{X}$. The code sequence $\underline{\boldsymbol{c}}$ is first interleaved by π. Next, the interleaved sequence $\pi(\underline{\boldsymbol{c}})$ is broken into subsequences of m bits each, which are mapped onto signals in $\mathcal{X}$. Finally, the resulting signal sequence $\underline{\boldsymbol{x}}$ is transmitted over the vector channel. The bit interleaver can be seen as a one-to-one correspondence $\pi : k \to (k', i)$, where k denotes the time ordering of the coded bits c_k, k' denotes the time ordering of the signals $\boldsymbol{x}_{k'}$ transmitted over the vector channel, and i indicates the position of the bit c_k in the label of $\boldsymbol{x}_{k'}$. Here we assume ideal bit interleaving.

Let $\ell^i(\boldsymbol{x})$ denote the ith bit of the label of $\boldsymbol{x}$, and $\mathcal{X}_b^i$ the subset of all the signals $\boldsymbol{x} \in \mathcal{X}$ whose label has the value $b \in \{0,1\}$ in position i (i.e., such that $\ell^i(\boldsymbol{x}) = b$.) Further, let $\overline{b}$ denote the complement of b and $\boldsymbol{y}$ the channel output resulting from the transmission of $\boldsymbol{x}$.

The conditional pdf of $\boldsymbol{y}$, given $\ell^i(\boldsymbol{x}) = b \in \{0,1\}$ and $\boldsymbol{\theta}$, is

$$\begin{aligned} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \ell^i(\boldsymbol{x}) = b) &= \sum_{\boldsymbol{x} \in \mathcal{X}} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) P(\boldsymbol{x} \mid \ell^i(\boldsymbol{x}) = b) \\ &= 2^{-(m-1)} \sum_{\boldsymbol{x} \in \mathcal{X}_b^i} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}). \end{aligned} \tag{6}$$

The last expression is obtained by noting that

$$p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}, \ell^i(\boldsymbol{x}) = b) = p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})$$

and that

$$P(\boldsymbol{x} \mid \ell^i(\boldsymbol{x}) = b) = \begin{cases} 0, & \text{if } \ell^i(\boldsymbol{x}) = \overline{b} \\ 2^{-(m-1)}, & \text{if } \ell^i(\boldsymbol{x}) = b \end{cases}$$

where we assume uniform input distribution

$$P(b=0) = P(b=1) = 1/2.$$

Again, the optimum receiver depends on the available CSI. For both cases of ideal CSI and no CSI with ideal bit interleaving, as we consider in this paper, at each time k' the demodulator DEM produces the set of ML *bit* metrics

$$\lambda^i(\boldsymbol{y}_{k'}, b) = \begin{cases} \log \sum\limits_{\boldsymbol{x} \in \mathcal{X}_b^i} p_{\boldsymbol{\theta}_{k'}}(\boldsymbol{y}_{k'} \mid \boldsymbol{x}), & \text{perfect CSI} \\ \log \sum\limits_{\boldsymbol{x} \in \mathcal{X}_b^i} p(\boldsymbol{y}_{k'} \mid \boldsymbol{x}), & \text{no CSI} \end{cases} \tag{7}$$

for $b \in \{0,1\}$ and $i = 1, \cdots, m$. Finally, the ML decoder DEC makes decisions according to the rule

$$\hat{\underline{\boldsymbol{c}}} = \arg\max_{\underline{\boldsymbol{c}} \in \mathcal{C}} \sum_k \lambda^i(\boldsymbol{y}_{k'}, c_k). \tag{8}$$

The decision rule (8) is completely analogous to (4), and can be implemented by a Viterbi decoder.

1) Simplified Branch Metrics: The branch metrics (7) may be computationally too complex for implementation. Suboptimal simplified branch metrics [26] can be obtained by the log-sum approximation $\log \sum_j z_j \simeq \max_j \log z_j$, which is good as long as the sum in the left-hand side is dominated by a single term, as typically occurs in channels with high signal-to-noise ratio (SNR). This yields

$$\tilde{\lambda}^i(\boldsymbol{y}_{k'}, b) = \begin{cases} \max\limits_{\boldsymbol{x} \in \mathcal{X}_b^i} \log p_{\boldsymbol{\theta}_{k'}}(\boldsymbol{y}_{k'} \mid \boldsymbol{x}), & \text{perfect CSI} \\ \max\limits_{\boldsymbol{x} \in \mathcal{X}_b^i} \log p(\boldsymbol{y}_{k'} \mid \boldsymbol{x}), & \text{no CSI} \end{cases} \tag{9}$$

Branch metrics (9) are obtained from the set of symbol branch metrics (3) by selecting the maximum over the subsets $\mathcal{X}_b^i$. BICM branch metric computation can be implemented by the same demodulator for CM followed by some post-processor (namely, by a parallel bank of comparators). Hence, BICM metric computation has the same latency of CM, but requires $m(2^m - 2)$ additional binary comparisons.

2) Gray Labeling: Gray labeling plays a key role in BICM theory. For convenience, we restate here its definition in terms of BICM notation.

Definition. Gray Labeling: Let $\mathcal{X}$ denote a signal set of size $M = 2^m$, with minimum Euclidean distance $d_{\min}$. A binary map $\mu : \{0,1\}^m \to \mathcal{X}$ is a Gray labeling for $\mathcal{X}$ if, for all $i = 1, \cdots, m$ and $b \in \{0,1\}$, each $\boldsymbol{x} \in \mathcal{X}_b^i$ has at most one $\boldsymbol{z} \in \mathcal{X}_{\overline{b}}^i$ at distance $d_{\min}$.

Fig. 2(b) shows a Gray-labeled 16QAM signal set.

3) Equivalent Channel Model: Fig. 3 shows the model of a BICM communication channel with ideal interleaving. The system consists of a set of m parallel *independent* and *memoryless* binary input channels connected to the encoder output by a random switch, which models ideal interleaving. Each channel corresponds to a position in the label of the signals of $\mathcal{X}$. For every symbol c_k of a coded sequence $\underline{\boldsymbol{c}}$, the switch selects randomly (and independently of other selections) a position index $i \in \{1, \cdots, m\}$, and transmits c_k on the ith channel. The decoder, which knows the sequence of channels used for

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

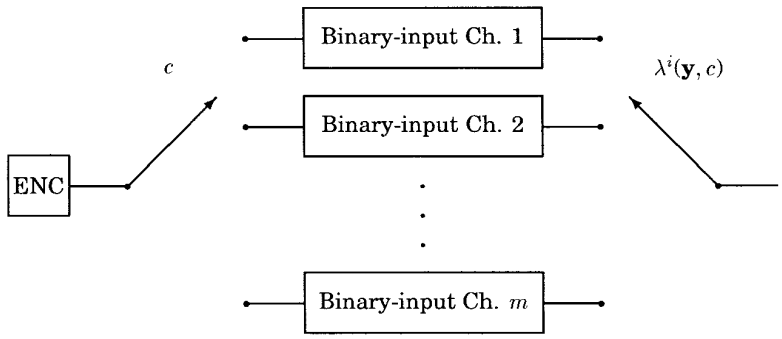

Fig. 3. Equivalent parallel channel model for BICM in the case of ideal interleaving.

the transmission of $\underline{c}$, computes the branch metrics (7) (or (9)) and selects $\hat{\underline{c}}$ according to (8).

Example 2. Zehavi's 8PSK Code: The effectiveness of BICM schemes over the Rayleigh fading channel was first pointed out by Zehavi in [26]. He picked for $\mathcal{C}$ the best 8-state, rate-$2/3$ binary convolutional code [11], and for μ a Gray-labeling map. Unlike with our BICM model, the output bits from the convolutional encoder were passed through three separate ideal interleavers, so that there was a fixed correspondence between the output bits of the encoder and the label positions. Apparently, there are no reasons justifying this fixed correspondence which, on the other hand, limits the flexibility of BICM and complicates the analysis. Moreover, a fixed correspondence between coded bits and label positions introduces unequal error protection, usually an undesired feature, and suboptimal performance when $\mathcal{C}$ is chosen at random. This motivates our BICM model: here all the coded bits output by the encoder are fed into a single-bit interleaver, and thus being on an equal footing, generate a maximum of symmetry.

III. An Information-Theoretical View of BICM

In this section we compute the capacity and the cutoff rate of CM and BICM with ideal interleaving, under the constraint of uniform input probabilities. As usual, we assume ideal interleaving, so that the sequence of channel state parameters $\boldsymbol{\theta}$ is i.i.d.

A. Channel Capacity

Consider the memoryless discrete-input, continuous-output channel with input $\boldsymbol{x}$, output $\boldsymbol{y}$, and transition distribution $p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})$. Under the assumption $I(\boldsymbol{x};\boldsymbol{\theta}) = 0$, from the "chain rule" of mutual information [10] we obtain the inequality $I(\boldsymbol{x};\boldsymbol{y}) \le I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{\theta})$. The right-hand- and left-hand-side terms of this inequality are attained when the receiver has perfect CSI and no CSI, respectively. Then, the capacity under uniform inputs constraint and perfect CSI is given by the conditional average mutual information (AMI)

$$\tilde{C} = I(\boldsymbol{x};\boldsymbol{y} \mid \boldsymbol{\theta}) = m - E_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta}}\left[\log_2 \frac{\sum_{\boldsymbol{z}\in\mathcal{X}} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}{p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})}\right] \tag{10}$$

where $m = \log_2 M$. Similarly, without CSI we get

$$\tilde{C} = I(\boldsymbol{x};\boldsymbol{y}) = m - E_{\boldsymbol{x},\boldsymbol{y}}\left[\log_2 \frac{\sum_{\boldsymbol{z}\in\mathcal{X}} p(\boldsymbol{y} \mid \boldsymbol{z})}{p(\boldsymbol{y} \mid \boldsymbol{x})}\right]. \tag{11}$$

Here, capacity is expressed in information bits per N complex dimensions (bit/dim). CM schemes, being signal space codes, i.e., codes whose words are sequences of signals $\boldsymbol{x} \in \mathcal{X}$, can achieve spectral efficiencies $R \le \tilde{C}$. For this reason, we will refer to $\tilde{C}$ defined in (10) as *CM capacity*.

Let us now compute the capacity achievable by BICM. To do it, we use the parallel-channel model of Fig. 3. Since the channels are memoryless and independent, we drop the time index k. Let b denote a binary input, $\boldsymbol{y}$ the vector channel output, and S the random variable whose outcome determines the switch position (for the above, S is i.i.d., uniformly distributed over $\{1, \cdots, m\}$, and known to the receiver). Since S, $\boldsymbol{\theta}$, and b are independent, we can show that the AMI of b and $\boldsymbol{y}$ satisfies the inequality $I(b;\boldsymbol{y} \mid S) \le I(b;\boldsymbol{y} \mid \boldsymbol{\theta}, S)$. Again, these mutual informations are attained with no CSI and with perfect CSI, respectively. The conditional mutual information of b and $\boldsymbol{y}$, given $S = i$, is then given by

$$I(b;\boldsymbol{y} \mid \boldsymbol{\theta}, S = i) = 1 - E_{b,\boldsymbol{y},\boldsymbol{\theta}}\left[\log_2 \frac{\sum_{\boldsymbol{z}\in\mathcal{X}} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}{\sum_{\boldsymbol{z}\in\mathcal{X}_b^i} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}\right] \tag{12}$$

where b, $\boldsymbol{y}$, and $\boldsymbol{\theta}$ are conditionally jointly distributed as

$$p(b, \boldsymbol{y}, \boldsymbol{\theta} \mid S = i) = 2^{-m} \sum_{\boldsymbol{z}\in\mathcal{X}_b^i} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z}) p_{\boldsymbol{\theta}}(\boldsymbol{\theta}). \tag{13}$$

The conditional AMI is obtained by averaging (12) with respect to S, so that

$$I(b;\boldsymbol{y} \mid \boldsymbol{\theta}, S) = \frac{1}{m}\sum_{i=1}^{m} I(b;\boldsymbol{y} \mid \boldsymbol{\theta}, S = i).$$

Finally, since there are m parallel independent channels, the BICM capacity, with perfect CSI and uniform inputs, is given by

$$\begin{aligned}\hat{C} &= mI(b;\boldsymbol{y} \mid \boldsymbol{\theta}, S)\\ &= m - \sum_{i=1}^{m} E_{b,\boldsymbol{y},\boldsymbol{\theta}}\left[\log_2 \frac{\sum_{\boldsymbol{z}\in\mathcal{X}} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}{\sum_{\boldsymbol{z}\in\mathcal{X}_b^i} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}\right].\end{aligned} \tag{14}$$

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

Similarly, with no CSI we obtain

$$\hat{C} = m - \sum_{i=1}^{m} E_{b,\boldsymbol{y}}\left[\log_2 \frac{\sum_{\boldsymbol{z}\in\mathcal{X}} p(\boldsymbol{y} \mid \boldsymbol{z})}{\sum_{\boldsymbol{z}\in\mathcal{X}_b^i} p(\boldsymbol{y} \mid \boldsymbol{z})}\right]. \tag{15}$$

In general, expectations in (10), (11), (14), and (15) cannot be calculated in closed form. Thus we resorted to numerical integration via the Monte Carlo method.

The data-processing theorem [10, p. 32] can finally be invoked to prove that, since $\boldsymbol{y}$ is statistically independent of b given $\boldsymbol{x}$ for all $\boldsymbol{\theta}$ and S, we have the inequality

$$\hat{C} \leq \tilde{C}. \tag{16}$$

This shows the suboptimality of BICM schemes.

B. Channel Cutoff Rate

Under the usual assumption of ideal interleaving and uniform input distribution, we now compute the cutoff rate of CM and BICM channels. This is widely accepted to be a sensible parameter for comparing channels for which a finite-complexity coding scheme is used.

The cutoff rate $\tilde{R}_0$ of the discrete-input continuous-output channel generated by a CM scheme can be written as [23], [14]

$$\tilde{R}_0 = \begin{cases} m - \log_2 \sum_{\boldsymbol{z}\in\mathcal{X}} E_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta}}\left[\sqrt{\frac{p_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{z})}{p_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}}\right], & \text{perfect CSI} \\ m - \log_2 \sum_{\boldsymbol{z}\in\mathcal{X}} E_{\boldsymbol{x},\boldsymbol{y}}\left[\sqrt{\frac{p(\boldsymbol{y}|\boldsymbol{z})}{p(\boldsymbol{y}|\boldsymbol{x})}}\right], & \text{no CSI.} \end{cases} \tag{17}$$

The cutoff rate of a BICM scheme can be obtained from the Bhattacharyya bound [23] on the average bit-error probability P_b of the parallel channel model of Fig. 3 in the absence of coding. By considering the ML bit metrics with perfect CSI given in (7), we obtain

$$P_b \leq \frac{1}{m}\sum_{i=1}^{m} E_{b,\boldsymbol{y},\boldsymbol{\theta}}\left[\sqrt{\frac{p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \bar{b}, i)}{p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid b, i)}}\right] = B \tag{18}$$

where B denotes the average Bhattacharyya factor of the BICM channel, with perfect CSI. Hence, we obtain from (6)

$$B \triangleq \frac{1}{m}\sum_{i=1}^{m} E_{b,\boldsymbol{y},\boldsymbol{\theta}}\left[\sqrt{\frac{\sum_{\boldsymbol{z}\in\mathcal{X}_{\bar{b}}^i} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}{\sum_{\boldsymbol{z}\in\mathcal{X}_b^i} p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z})}}\right]. \tag{19}$$

Similarly, by considering the ML bit metrics without CSI given in (7), we get the average Bhattacharyya factor

$$B \triangleq \frac{1}{m}\sum_{i=1}^{m} E_{b,\boldsymbol{y}}\left[\sqrt{\frac{\sum_{\boldsymbol{z}\in\mathcal{X}_{\bar{b}}^i} p(\boldsymbol{y} \mid \boldsymbol{z})}{\sum_{\boldsymbol{z}\in\mathcal{X}_b^i} p(\boldsymbol{y} \mid \boldsymbol{z})}}\right]. \tag{20}$$

Finally, note that a channel use of the BICM channel is equivalent to m channel uses of a binary-input channel with average Bhattacharyya factor B. Hence, the cutoff rate $\hat{R}_0$ of BICM can be written as

$$\hat{R}_0 = m(1 - \log_2(B+1)) \tag{21}$$

where B is given by (19) with perfect CSI or by (20) with no CSI. In spite of the fact that $\tilde{R}_0$ can be evaluated analytically in many cases [14], B, and hence $\hat{R}_0$, cannot. Again, as we did for capacity, we resorted to Monte Carlo numerical integration to calculate $\hat{R}_0$.

C. Numerical Calculations for Fading Channels

While the results obtained in the previous section hold in general, hereafter we restrict our attention the slow frequency nonselective Rician fading channels [14], which encompass Rayleigh and AWGN as special cases. With this model, the received signal after the transmission of $\boldsymbol{x} \in \mathcal{X}$ can be written as

$$\boldsymbol{y} = g\boldsymbol{x}e^{j\phi} + \boldsymbol{n} \tag{22}$$

where: i) $\boldsymbol{n}$ is a complex zero-mean Gaussian i.i.d. random vector with covariance $\frac{1}{2}\mathrm{E}[\boldsymbol{n}\boldsymbol{n}^\dagger] = \frac{1}{2}\sigma^2\boldsymbol{I}_N$ ($\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and a dagger $\dagger$ denotes Hermitian transpose); ii) g is a scalar complex fading gain;[2] iii) ϕ is the carrier phase, independent of $\boldsymbol{x}$ and g and uniformly distributed over $[-\pi, \pi]$.

With Rician fading, g is complex Gaussian with mean $\sqrt{K/(K+1)}$ and variance $1/(2(K+1))$ (where $K \geq 0$ is the *Rician parameter*, defined as the ratio of direct to diffuse power in the received signal [14]). We assume normalized signal energy $\frac{1}{2}E[|\boldsymbol{x}|^2] = 1/2$, so that the average SNR is $1/\sigma^2$. With this model $\boldsymbol{\theta} = (g, \phi)$ and $p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})$ is given by

$$p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{1}{(\pi\sigma^2)^N}\exp\left(-|\boldsymbol{y} - g\boldsymbol{x}e^{j\phi}|^2/\sigma^2\right). \tag{23}$$

By averaging with respect to g and ϕ, we obtain [12]

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{x}) &= E_{\boldsymbol{\theta}}[p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})] \\ &= \frac{1}{\pi^N|\Sigma_y|}\exp\left(-\boldsymbol{y}^\dagger\Sigma_y^{-1}\boldsymbol{y} - \bar{\boldsymbol{y}}^\dagger\Sigma_y^{-1}\bar{\boldsymbol{y}}\right)I_0\left(2\left|\bar{\boldsymbol{y}}^\dagger\Sigma_y^{-1}\boldsymbol{y}\right|\right) \end{aligned} \tag{24}$$

where we have defined

$$\bar{\boldsymbol{y}} = \sqrt{\frac{K}{K+1}}\boldsymbol{x}, \quad \text{and} \quad \Sigma_y = \frac{1}{K+1}\boldsymbol{x}\boldsymbol{x}^\dagger + \sigma^2\boldsymbol{I}_N$$

and where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero [17].

Numerical results can be obtained by using (23) and (24) in the general expressions of $\tilde{C}$, $\hat{C}$, $\tilde{R}_0$, and $\hat{R}_0$, for any choice of signal set and labeling map. Here we restrict ourselves to the limiting cases of AWGN and Rayleigh fading channels ($K \to \infty$ and $K = 0$, respectively). While looking at the following results, the reader should keep in mind that *bit/dim* indicates information bits per *complex dimension*.

1) PSK and QAM with Perfect CSI: Figs. 4 and 5 show CM and BICM capacity versus SNR for 4PSK, 8PSK, and 16QAM over the AWGN and Rayleigh fading channels, respectively. BICM capacity is computed with Gray and set partitioning (SP) labelings. The performance of BICM with Gray labeling is very close to that of CM (actually,

[2] This model does not apply when component interleaving is used in addition to bit interleaving, since in this case the N components of $\boldsymbol{x}$ are affected by different fading coefficients [6], [1].

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

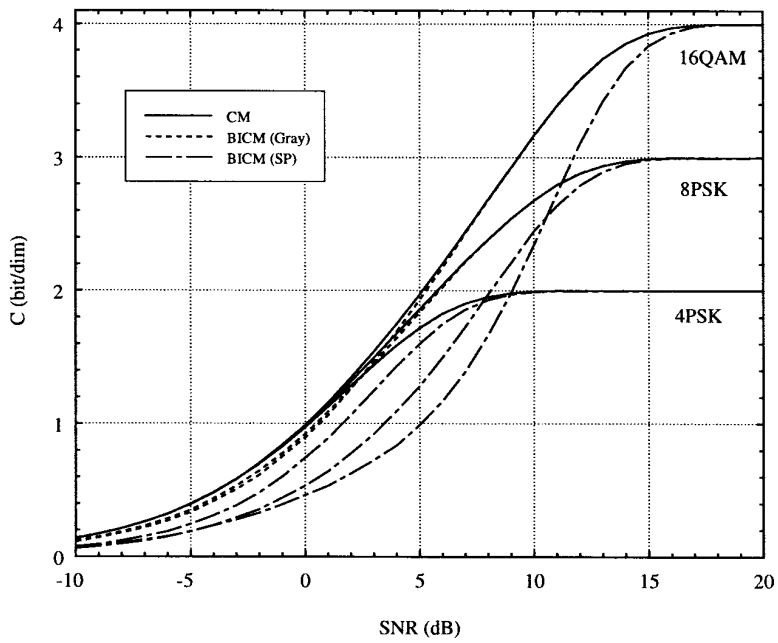

Fig. 4. BICM and CM capacity versus. SNR for 4PSK, 8PSK, and 16QAM over AWGN with coherent detection (SP denotesset-partitioning labeling).

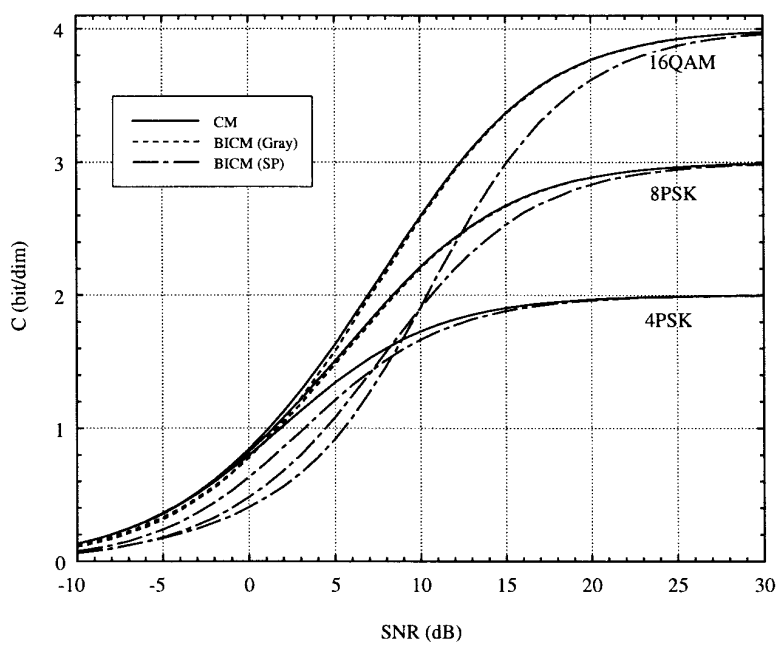

Fig. 5. BICM and CM capacity versus SNR for 4PSK, 8PSK, and 16QAM over Rayleigh fading with coherent detection and perfect CSI (SP denotes set-partitioning labeling).

we conjecture that Gray labeling maximizes BICM capacity) while the performance of BICM with SP labeling is several decibels worse. Similar differences between Gray and SP labelings can also be observed from the cutoff rate of the Rayleigh fading channel.

Our next results are based on cutoff rate. This parameter appears to be more suitable than capacity to compare BICM and CM, possibly because there is no fixed relation between $\hat{R}_0$ and $\tilde{R}_0$ whereas $\hat{C} \leq \tilde{C}$, as we know from (16). Figs. 6 and 7 show BICM and CM cutoff rate versus SNR for QAM signal sets, over the AWGN and Rayleigh fading channels, respectively. For 4, 16, 64, and 256QAM signal sets we used Gray labeling. For 8, 32, and 128QAM—for which Gray labeling is not possible—we used a *quasi*-Gray labeling, i.e., a labeling minimizing the number of signals for which the Gray condition (of having at most one nearest neighbor in the complement subset) is not satisfied. For AWGN, CM outperforms BICM at all SNR. The performance gap, which is large for large M and low-rate codes, is reduced for high-rate codes. For example, CM256 QAM gains more than 3

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

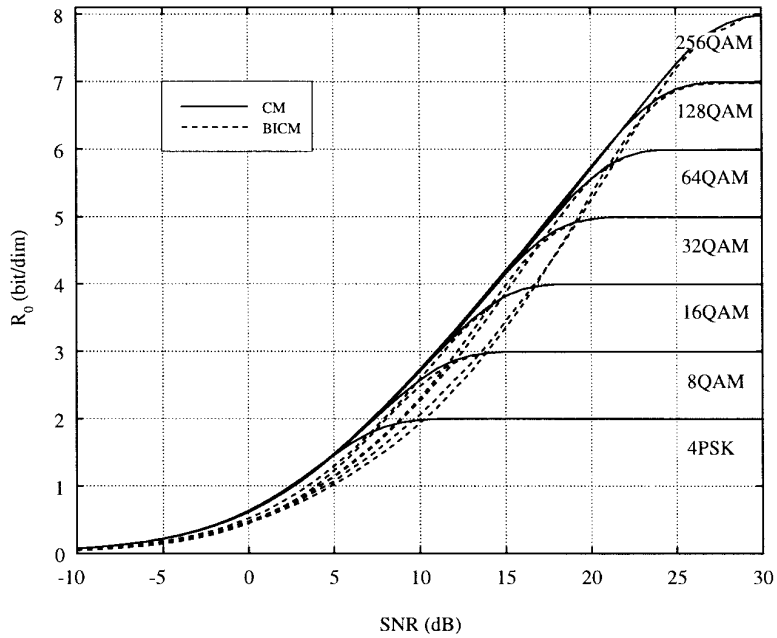

Fig. 6. BICM and CM cutoff rate versus SNR for QAM signal sets with Gray (or quasi-Gray) labeling over AWGN with coherent detection.

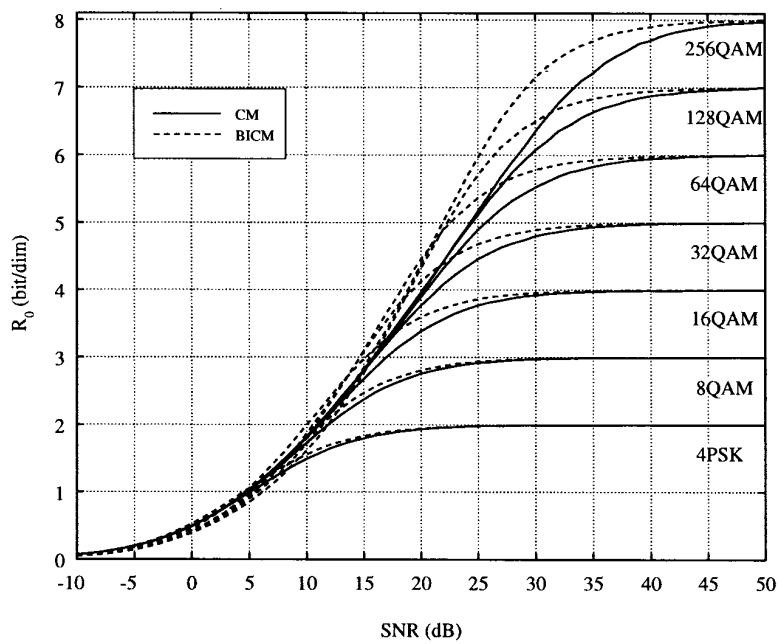

Fig. 7. BICM and CM cutoff rate versus SNR for QAM signal sets with Gray (or quasi-Gray) labeling over Rayleigh fading with coherent detection and perfect CSI.

dB over BICM 256QAM at $R_0 = 3$ bit/dim, but this gap is reduced to less than 0.5 dB at $R_0 = 7$ bit/dim. The situation is different for the Rayleigh channel. We observe that in this case, BICM generally outperforms CM for $R_0 > 1$ bit/dim. This difference in performance is especially apparent for high-rate codes: for example, BICM 256QAM gains about 4 dB over CM 256QAM at $R_0 = 7$ bit/dim. The above fact can be intuitively explained as follows. For low rates it is possible to design "practical" CM codes with large M-ary Hamming distance, so that their performance over the Rayleigh fading channel is good. For high rates, the complexity required by CM in order to obtain large M-ary Hamming distance is much larger than the complexity required by BICM, for which Hamming distance is given by the binary Hamming distance of the underlying binary code. Hence, for a given complexity, BICM compares favorably with respect to CM on the Rayleigh fading channel, especially for high rates.

Capacity and cutoff rate curves provide guidelines for code design. In particular, over the AWGN channel CM appears more suitable, although for high rates the loss of optimality

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

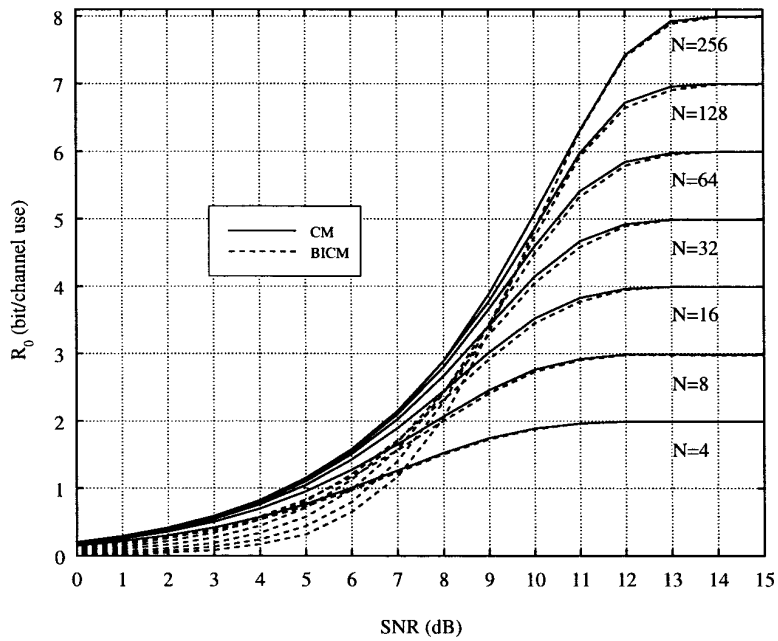

Fig. 8. BICM and CM cutoff rate versus. SNR for orthogonal signal sets over AWGN with noncoherent detection. The cutoff rate is expressed in information bit per channel use, where a channel use corresponds to N complex dimensions.

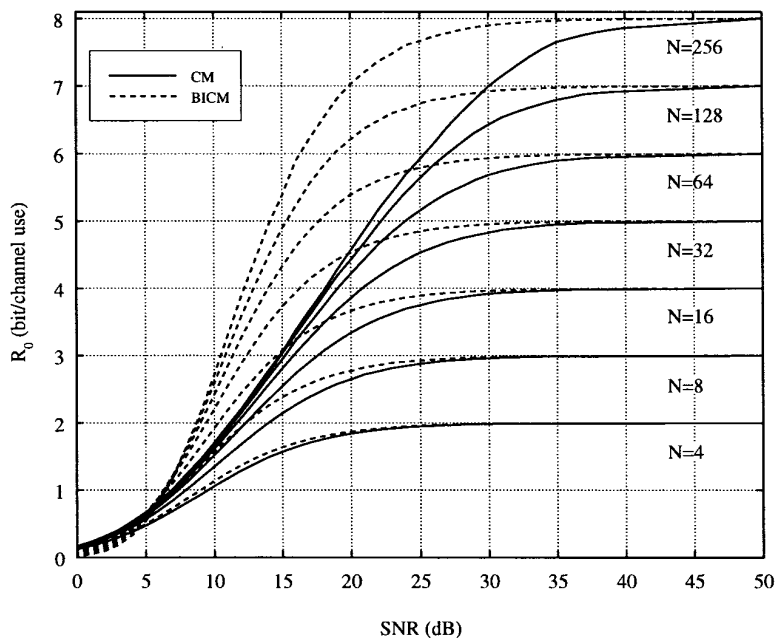

Fig. 9. BICM and CM cutoff rate versus SNR for orthogonal signal sets over Rayleigh fading with noncoherent detection and no CSI. The cutoff rate is expressed in information bit per channel use, where a channel use corresponds to N complex dimensions.

introduced by BICM is marginal (thus leaving room for "pragmatic" approaches [24]). On the contrary, BICM is much more appropriate for the Rayleigh fading channel. As a consequence, if the channel model—as is the case for example for mobile radio—fluctuates in time between the extremes of Rayleigh and AWGN, BICM proves to be a more robust choice than CM.

2) N-ary Orthogonal Signals with No CSI: We consider a unit-energy N-ary orthogonal signal set $\mathcal{X}$ with $N = 2^m$ (the sequences $\underline{x} \in \mathcal{X}$ can be obtained, for example, as unit-energy Hadamard sequences [22]). Detection is noncoherent with no CSI. Given the symmetry of orthogonal signals, $\hat{C}$ and $\hat{R}_0$ do not depend on the labeling μ, which can be arbitrarily chosen.

Again, we use the cutoff rate to compare BICM and CM. Figs. 8 and 9 show BICM and CM cutoff rates for $N = 4, 8, 16, 32, 64, 128$, and 256 over AWGN and Rayleigh fading channels, respectively. An application might be coded DS/SSMA, for which orthogonal signals are used to obtain

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

very low rate CM schemes.[3] At $R_0 = 1/N$ bit/dim, CM outperforms BICM over the AWGN channel, especially for large N. On the contrary, BICM outperforms CM over the Rayleigh fading channel. We note from Fig. 9 that, for the Rayleigh fading channel, BICM provides the largest gain over CM for spectral efficiencies higher than $1/N$ (for example, at $R_0 = 7/256$ bit/dim, BICM yields a 10-dB gain over CM). However, efficiencies better than $1/N$ are not very interesting for coded DS/SSMA applications, where signal bandwidth expansion is obtained by low-rate coding. In Section V we shall see that actual BICM schemes can provide a substantial coding gain over actual CM schemes with the same complexity also at rate $1/N$ bit/dim.

IV. Error-Probability Analysis

In this section we derive tools for evaluating P_b, the bit error rate (BER) of BICM with ideal interleaving. From now on we assume that the binary code $\mathcal{C}$ in the BICM model is linear, so that it admits a (possibly time-varying) minimal trellis representation [18].

Let $\underline{\boldsymbol{c}}$ and $\underline{\hat{\boldsymbol{c}}}$ denote two distinct sequences stemming from the same state and merging after $\ell \geq 1$ trellis steps. Denote by $\{\underline{\boldsymbol{c}} \to \underline{\hat{\boldsymbol{c}}}\}$ a decoder *pairwise error event* (i.e., the decoder chooses $\underline{\hat{\boldsymbol{c}}}$ when $\underline{\boldsymbol{c}}$ is the transmitted sequence and $\underline{\boldsymbol{c}}$ and $\underline{\hat{\boldsymbol{c}}}$ are the only two possible decoding outcomes) and by $P(\underline{\boldsymbol{c}} \to \underline{\hat{\boldsymbol{c}}})$ the *pairwise error probability* (PEP). The binary-input channels of the BICM parallel channel model of Fig. 3 may be nonsymmetric, depending on the labeling map μ and on the signal set $\mathcal{X}$. Then, even for a linear $\mathcal{C}$, the PEP may depend on the pair $(\underline{\boldsymbol{c}}, \underline{\hat{\boldsymbol{c}}})$ rather than on their difference $\underline{\boldsymbol{c}} \oplus \underline{\hat{\boldsymbol{c}}}$. The analysis can be made simpler if we symmetrize the BICM parallel-channel model by introducing a time-varying labeling map. Define the complement $\bar{\mu}$ of the map μ as the labeling obtained by complementing all the signal labels. Further, let U denote a binary random variable independent of S, b, and $\boldsymbol{\theta}$, which for each transmission over the vector channel selects either μ or $\bar{\mu}$, independently with probability $1/2$. Then, any binary symbol b, input to the parallel-channel model, has probability $1/2$ of being complemented before transmission (which occurs when $U = 1$), so that the BICM channel is made symmetric. We assume that the value taken on by U is known to the receiver, so that further complementing restores the actual value of b. Under the common hypothesis of uniform input bits, performance is not degraded by symmetrization. The PEP of symmetrized BICM, with ideal interleaving, is a function of the Hamming distance d between $\underline{\boldsymbol{c}}$ and $\underline{\hat{\boldsymbol{c}}}$, of the labeling μ and of the signal set $\mathcal{X}$. Therefore, we can write

$$P(\underline{\boldsymbol{c}} \to \underline{\hat{\boldsymbol{c}}}) = f(d, \mu, \mathcal{X}) \tag{25}$$

where the function $f(\cdot,\cdot,\cdot)$ depends on the vector channel and on the type of detection. Without loss of generality, in the following we assume that $\underline{\boldsymbol{c}}$ and $\underline{\hat{\boldsymbol{c}}}$ differ in the d consecutive positions $1, \cdots, d$. Finally, the usual union bound for linear binary codes leads to

$$\begin{cases} P_\mathrm{b} \leq \frac{1}{k_c} \sum\limits_{d=1}^{\infty} W_I(d) f(d, \mu, \mathcal{X}) \\ \qquad\qquad \text{convolutional codes of rate } k_c/n_c \\ P_\mathrm{b} \leq \frac{1}{K_c} \sum\limits_{d=1}^{N_c} W_I(d) f(d, \mu, \mathcal{X}) \\ \qquad\qquad \text{block codes of rate } K_c/N_c \end{cases} \tag{26}$$

where $W_I(d)$ denotes the total input weight of error events at Hamming distance d. With convolutional codes, the infinite summation of the union bound is assumed to converge, for sufficiently large SNR. In the next section we shall see that this assumption holds, in general, for the BICM model with ideal interleaving adopted in this paper and for well-behaved channels. Methods for computing the coefficients $W_I(d)$ are well known, especially for convolutional codes (see, e.g., [23].) Thus the key quantity to be evaluated in the following is the PEP $f(d, \mu, \mathcal{X})$.

A. Bhattacharyya Union Bound

Let $\underline{\boldsymbol{S}}$, $\underline{\boldsymbol{U}}$, and $\underline{\boldsymbol{\theta}}$ denote the sequences of label positions, labeling maps, and channel states occurring during the transmission of $\underline{\boldsymbol{c}}$. A standard upper bound on the PEP is through the Bhattacharyya bound [23]. For an ML decoder (8) with perfect CSI we get

$$\begin{aligned} & f(d, \mu, \mathcal{X}) \\ & \quad = E_{\underline{\boldsymbol{S}},\underline{\boldsymbol{U}}}[P(\underline{\boldsymbol{c}} \to \underline{\hat{\boldsymbol{c}}} \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}})] \\ & \quad = E_{\underline{\boldsymbol{S}},\underline{\boldsymbol{U}}}[P(p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{c}}, \underline{\boldsymbol{S}}) \leq p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\hat{\boldsymbol{c}}}, \underline{\boldsymbol{S}}) \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}})] \\ & \quad \leq E_{\underline{\boldsymbol{S}},\underline{\boldsymbol{U}}}\left[E_{\underline{\boldsymbol{y}},\underline{\boldsymbol{\theta}}}\left[\sqrt{\frac{p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\hat{\boldsymbol{c}}}, \underline{\boldsymbol{S}})}{p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{c}}, \underline{\boldsymbol{S}})}} \,\middle|\, \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}}\right]\right] \\ & \quad = \prod_{k=1}^{d} \frac{1}{m} \sum_{i_k=1}^{m} E_{U_k}\left[E_{\boldsymbol{y}_k,\boldsymbol{\theta}_k}\left[\sqrt{\frac{p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid \hat{c}_k, i_k)}{p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid c_k, i_k)}} \,\middle|\, U_k\right]\right] \\ & \quad = \left(\frac{1}{m} \sum_{i=1}^{m} E_{b,\boldsymbol{y},\boldsymbol{\theta}}\left[\sqrt{\frac{p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \bar{b}, i)}{p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid b, i)}}\right]\right)^d = B^d. \end{aligned} \tag{27}$$

In a similar way, without CSI we obtain $f(d, \mu, \mathcal{X}) \leq B^d$, with B given by (20).

For 8PSK with Gray labeling, perfect CSI, and the modified metric (9), a Chernoff bound on the PEP was obtained by Zehavi [26, Appendix A]. Although the latter can be expressed in closed form, its derivation relies upon the specific features of 8PSK and Gray labeling, and cannot be applied directly to other labelings and/or other signal sets. By using (27) in (26), on the contrary, we obtain bounds on P_b for any labeling, signal set, and vector channel, with and without CSI. The drawback here is that we must evaluate numerically B: however, numerical evaluation is much faster than simulation of the whole BICM system, and allows evaluation of P_b for values usually not reachable by simulation. Moreover, B is independent of $\mathcal{C}$, and hence it can be evaluated for a given μ and $\mathcal{X}$, and used to compare the performances of different codes on any given channel.

[3] See [22]. Viterbi's orthogonal convolutional codes are in fact CM codes, as explained in Section V.

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

Convergence of the Union Bound: Under the assumptions of Section II-A, with ideal interleaving the channel is memoryless and stationary. Hence, the random coding bound based on the Bhattacharyya union bound for (time-varying) convolutional codes given in [23, Ch. 5] can be applied to our case. In particular, for $R < \hat{R}_0$ the bound converges. For actual codes, by replacing $f(d, \mu, \mathcal{X})$ with B^d in the infinite sum of (26), we easily see that there exists a B_0 such that if $B < B_0$ the bound converges. Then, as long as B is a decreasing function of the channel SNR, we can safely assume that the union bound converges over a range of SNR values above a certain threshold.

B. BICM Union Bound

In this subsection we derive a union bound which does not rely on Bhattacharyya (or Chernoff) upper bounds. This turns out to be quite loose, but we shall use it only as a starting point for the tight bounds and approximations which will be developed in the next section. Since the bit metrics (7) or (9) are difficult to deal with, our technique consists of going back to the original vector channel and finding expressions in terms of PEP's of pairs of signal sequences, input to the vector channel underlying the BICM channel.

In the following we consider the modified metrics (9) and perfect CSI. Assume that $\underline{\boldsymbol{c}}$ and $\hat{\underline{\boldsymbol{c}}}$ have Hamming distance d, and let

$$\begin{aligned} \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}} &= \mathcal{X}_{c_1}^{i_1} \times \cdots \times \mathcal{X}_{c_k}^{i_k} \times \cdots \times \mathcal{X}_{c_d}^{i_d} \\ \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}} &= \mathcal{X}_{\hat{c}_1}^{i_1} \times \cdots \times \mathcal{X}_{\hat{c}_k}^{i_k} \times \cdots \times \mathcal{X}_{\hat{c}_d}^{i_d} \end{aligned} \tag{28}$$

denote the sequence (Cartesian product) of signal subsets selected by the bits c_k of $\underline{\boldsymbol{c}}$ (correct subsets) and $\hat{c}_k$ of $\hat{\underline{\boldsymbol{c}}}$ (complement subsets) given the label positions $S_k = i_k$ and the labeling map selected by U_k. For given $\underline{\boldsymbol{S}}$, $\underline{\boldsymbol{U}}$, and $\underline{\boldsymbol{\theta}}$, the path metric difference corresponding to the sequences $\underline{\boldsymbol{c}}$ and $\hat{\underline{\boldsymbol{c}}}$ can be written as

$$\begin{aligned} \delta &= \sum_{k=1}^{d} \max_{\boldsymbol{z}_k \in \mathcal{X}_{c_k}^{i_k}} \log p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid \boldsymbol{z}_k) - \sum_{k=1}^{d} \max_{\boldsymbol{z}_k \in \mathcal{X}_{\hat{c}_k}^{i_k}} \log p_{\boldsymbol{\theta}_k}(\boldsymbol{y}_k \mid \boldsymbol{z}_k) \\ &= \max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \log p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) - \max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} \log p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}). \end{aligned} \tag{29}$$

This yields

$$\begin{aligned} &P(\underline{\boldsymbol{c}} \to \hat{\underline{\boldsymbol{c}}} \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}}) \\ &\quad = P(\delta \le 0 \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}}) \\ &\quad = P\left(\max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \le \max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}} \right) \\ &\quad \overset{\text{a)}}{=} E_{\underline{\boldsymbol{x}}} \left[P\left(\max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \le \max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}}, \underline{\boldsymbol{x}} \right) \right] \\ &\quad \overset{\text{b)}}{\le} E_{\underline{\boldsymbol{x}}} \left[P\left(p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}) \le \max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \mid \underline{\boldsymbol{S}}, \underline{\boldsymbol{U}}, \underline{\boldsymbol{x}} \right) \right] \\ &\quad \overset{\text{c)}}{\le} E_{\underline{\boldsymbol{x}}} \left[\sum_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} P(p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}) \le p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \mid \underline{\boldsymbol{x}}) \right] \\ &\quad \overset{\text{d)}}{=} 2^{-d(m-1)} \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \sum_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} P(\underline{\boldsymbol{x}} \to \underline{\boldsymbol{z}}) \end{aligned} \tag{30}$$

where a) is obtained by conditioning and averaging with respect to the transmitted signal sequence $\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$; b) is obtained from the obvious relation

$$p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}) \le \max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}});$$

c) is obtained by the union bound over all the sequences $\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$; and d) is obtained by assuming $\underline{\boldsymbol{x}}$ uniformly distributed over $\mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$ and by noting that

$$P(p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}) \le p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}}) \mid \underline{\boldsymbol{x}}) = P(\underline{\boldsymbol{x}} \to \underline{\boldsymbol{z}}) \tag{31}$$

is the PEP of the two signal sequences $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{z}}$, transmitted over the vector channel.

We are now ready to bound the PEP of BICM: it suffices to average (30) with respect to $\underline{\boldsymbol{S}}$ and $\underline{\boldsymbol{U}}$. We obtain

$$\begin{aligned} f(d, \mu, \mathcal{X}) &\le m^{-d} \sum_{\underline{\boldsymbol{S}}} 2^{-d} \sum_{\underline{\boldsymbol{U}}} 2^{-d(m-1)} \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \sum_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} P(\underline{\boldsymbol{x}} \to \underline{\boldsymbol{z}}) \\ &\triangleq f_{\text{ub}}(d, \mu, \mathcal{X}) \end{aligned} \tag{32}$$

where $\underline{\boldsymbol{S}}$ and $\underline{\boldsymbol{U}}$ run over all the possible m^d and 2^d sequences of position and labeling maps, respectively. We will refer to (32) as to the *BICM Union Bound* (UB).

The large number of terms in the right-hand-side summation makes (32) impractical, even for moderate values of d and m. For example, if $\mathcal{X}$ is 16QAM and $d = 10$ (as for the free distance of the best 64-state rate-1/2 binary convolutional code [17, pp. 466–471]) we should evaluate and sum up $4^{10} \times 2^{10} \times 2^{10(4-1)} \times 2^{10(4-1)} = 2^{90}$ terms. In order to get a computationally efficient form for the BICM UB, let

$$\Delta(\boldsymbol{x}, \boldsymbol{z}) = \log p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) - \log p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{z}) \tag{33}$$

denote the metric difference relative to the components $\boldsymbol{x}$ and $\boldsymbol{z}$ of the sequences $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{z}}$, and

$$\Delta = \sum_{k=1}^{d} \Delta(\boldsymbol{x}_k, \boldsymbol{z}_k)$$

denote the total path metric difference corresponding to $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{z}}$ (note that $\boldsymbol{x}_k \ne \boldsymbol{z}_k$ for all $k = 1, \cdots, d$, since $\mathcal{X}_{c_k}^{i_k} \cap \mathcal{X}_{\hat{c}_k}^{i_k} = \emptyset$). Then, we can write [17]

$$P(\underline{\boldsymbol{x}} \to \underline{\boldsymbol{z}}) = P(\Delta \le 0) = \frac{1}{2\pi j} \int_{\alpha - j\infty}^{\alpha + j\infty} \Phi_\Delta(s) \frac{ds}{s} \tag{34}$$

where $\Phi_\Delta(s)$ denotes the Laplace transform of the pdf of Δ, given by

$$\Phi_\Delta(s) = E[e^{-s\Delta}] = \prod_{k=1}^{d} E[e^{-s\Delta(\boldsymbol{x}_k, \boldsymbol{z}_k)}] = \prod_{k=1}^{d} \Phi_{\Delta(\boldsymbol{x}_k, \boldsymbol{z}_k)}(s) \tag{35}$$

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

and where $\alpha \in (0, \alpha_0)$, the intersection of the region of convergence of $\Phi_\Delta(s)$ with the positive real line. Finally, by using (34) into (32) and by exchanging the sums and the integral, we get the BICM UB in the form

$$f_{\rm ub}(d, \mu, \mathcal{X}) = \frac{1}{2\pi j} \int_{\alpha - j\infty}^{\alpha + j\infty} [\psi_{\rm ub}(s)]^d \, \frac{ds}{s} \tag{36}$$

where we have defined

$$\psi_{\rm ub}(s) \triangleq \frac{1}{m2^m} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{\boldsymbol{x} \in \mathcal{X}_b^i} \sum_{\boldsymbol{z} \in \mathcal{X}_{\bar{b}}^i} \Phi_{\Delta(\boldsymbol{x}, \boldsymbol{z})}(s) \tag{37}$$

and α belongs to the intersection of the regions of convergence of the $\Phi_{\Delta(\boldsymbol{x},\boldsymbol{z})}(s)$'s (for all pairs $\boldsymbol{x}, \boldsymbol{z}$) with the real positive line. Use of this technique is made easy by the fact that $\Phi_{\Delta(\boldsymbol{x},\boldsymbol{z})}(s)$ (and its convergence region) can be expressed in closed form in many cases of interest [3]. The integral over the vertical path $(\alpha - j\infty, \alpha + j\infty)$ can be efficiently computed by the method developed by the authors in [4] and based on a change of variable and Gauss–Chebyshev quadratures. Alternatively, one can use the looser bound described below (see [3] and references therein), which we state here without proof.

1) Chernoff Bound: Let Δ be a continuous real random variable which takes on values over the whole real line and let $\Phi_\Delta(s) = E[e^{-s\Delta}]$ denote the Laplace transform of the pdf of Δ. Then: i) The convergence region of $\Phi_\Delta(s)$ is the vertical strip $\alpha_1 < \mathrm{Re}(s) < \alpha_2$ where $\alpha_1 < 0 < \alpha_2$; ii) The following inequality holds:

$$P(\Delta \leq 0) \leq \min_{0 < \alpha < \alpha_2} \Phi_\Delta(\alpha). \tag{38}$$

Then, by noting that $[\psi_{\rm ub}(s)]^d$ is a weighted sum of Laplace transforms of the form (35), we obtain the Chernoff-type bound

$$f_{\rm ub}(d, \mu, \mathcal{X}) \leq \min_{0 < \alpha < \alpha_0} [\psi_{\rm ub}(\alpha)]^d. \tag{39}$$

We hasten to say that, unless the minimization with respect to α can be obtained in closed form, the computation of the right-hand side of (39) and (36) have roughly the same (low) complexity. Thus there is no real benefit in using the looser bound (39) instead of (36).

The BICM UB is very general, and can be easily extended to the case of no CSI by defining the appropriate path metric difference between signal sequences $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{z}}$ (we omit here this trivial derivation). Yet, it is rather loose, due to steps b) and c) of (30). Step b) is a strict inequality if

$$\underline{\boldsymbol{x}} \neq \arg\max_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \log p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{z}})$$

while step c) is a union bound which generates a large number of terms corresponding to almost overlapping pairwise-error events.

Example 3. Uncoded 4PSK in AWGN: Consider uncoded transmission of 4PSK with Gray labeling in AWGN with ideal coherent detection. Since 4PSK with Gray labeling provides two binary antipodal parallel independent channels, the presence of a bit interleaver is irrelevant, so that the exact BER is given by [17] $P_{\rm b} = f(1, \mu, \mathcal{X}) = Q(1/\sigma)$. By applying the BICM UB we get

$$P_{\rm b} \leq f_{\rm ub}(1, \mu, \mathcal{X}) = Q(1/\sigma) + Q(\sqrt{2}/\sigma). \tag{40}$$

The two terms on the right-hand side correspond to pairwise errors at distance $\sqrt{2}$ and 2, respectively. Hence, the BICM UB takes into account the contribution of events at distance 2, even if they are irrelevant to the BER. This simple observation calls for some "expurgation" technique in order to exclude the irrelevant terms from the BICM UB. This approach will be formally stated in the following section for AWGN and fading channels with perfect CSI.

C. Bounds for Fading Channels with Perfect CSI

In this subsection we derive a tight upper bound to the PEP of BICM for Rician fading channels with perfect CSI. Here $p_{\boldsymbol{\theta}}(\underline{\boldsymbol{y}} \mid \boldsymbol{x})$ takes the form (23), where, without loss of generality, we assume zero carrier phase. The bound is first derived for AWGN $(K \to \infty)$, and then extended to the fading case $(K < \infty)$.

With AWGN, maximizing $p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}})$ over the possible signal sequences is tantamount to minimizing the squared Euclidean distance $|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}|^2$. We exploit the signal space geometry to expurgate the BICM union bound. Let

$$\Gamma_{\underline{\boldsymbol{x}},\underline{\boldsymbol{z}}} = \{\underline{\boldsymbol{y}} \,:\, |\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}|^2 \geq |\underline{\boldsymbol{y}} - \underline{\boldsymbol{z}}|^2\} \tag{41}$$

denote the pairwise error region of the two signal sequences $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{z}}$, so that $P(\underline{\boldsymbol{x}} \to \underline{\boldsymbol{z}}) = P(\Gamma_{\underline{\boldsymbol{x}},\underline{\boldsymbol{z}}})$. Given $\underline{\boldsymbol{S}}$ and $\underline{\boldsymbol{U}}$, let $\hat{\underline{\boldsymbol{x}}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$ and $\hat{\underline{\boldsymbol{z}}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$ be two particular signal sequences corresponding to $\underline{\boldsymbol{c}}$ and $\hat{\underline{\boldsymbol{c}}}$, respectively. Then, we have

Theorem 1. Irrelevant Error Events (AWGN): The term $P(\hat{\underline{\boldsymbol{x}}} \to \hat{\underline{\boldsymbol{z}}})$ can be neglected in (30) without altering the inequality if there exist ℓ sequences $\underline{\boldsymbol{z}}^{(i)}$ different from $\hat{\underline{\boldsymbol{x}}}$ and $\hat{\underline{\boldsymbol{z}}}$, for all $i = 1, \cdots, \ell$, either in the complement subsets sequence $\mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$ or in the correct subsets sequence $\mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$, such that

$$\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} \subseteq \bigcup_{i=1}^{\ell} \Gamma_{\hat{\underline{\boldsymbol{x}}},\underline{\boldsymbol{z}}^{(i)}} \tag{42}$$

Proof: From (42) we write

$$\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} = \bigcup_{i=1}^{\ell} \{\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} \cap \Gamma_{\hat{\underline{\boldsymbol{x}}},\underline{\boldsymbol{z}}^{(i)}}\}.$$

Consider separately each subregion $\{\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} \cap \Gamma_{\hat{\underline{\boldsymbol{x}}},\underline{\boldsymbol{z}}^{(i)}}\}$. Trivially, if $\underline{\boldsymbol{z}}^{(i)} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$, $\{\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} \cap \Gamma_{\hat{\underline{\boldsymbol{x}}},\underline{\boldsymbol{z}}^{(i)}}\}$ is already counted in the pairwise error event $\{\hat{\underline{\boldsymbol{x}}} \to \underline{\boldsymbol{z}}^{(i)}\}$. The case $\underline{\boldsymbol{z}}^{(i)} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$ must be considered more carefully. In this case, for $\underline{\boldsymbol{y}} \in \{\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} \cap \Gamma_{\hat{\underline{\boldsymbol{x}}},\underline{\boldsymbol{z}}^{(i)}}\}$

$$\hat{\underline{\boldsymbol{x}}} \neq \arg\max_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \log p_{\underline{\boldsymbol{\theta}}}(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}).$$

Hence, the BICM decoder path metric for $\underline{\boldsymbol{c}}$ is not the metric corresponding to $\hat{\underline{\boldsymbol{x}}}$. Therefore, the subregion $\{\Gamma_{\hat{\underline{\boldsymbol{x}}},\hat{\underline{\boldsymbol{z}}}} \cap \Gamma_{\hat{\underline{\boldsymbol{x}}},\underline{\boldsymbol{z}}^{(i)}}\}$

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

does not belong to the pairwise error region of the BICM decoder, and can be neglected. We conclude that $\Gamma_{\underline{\hat{x}},\underline{\hat{z}}}$ is irrelevant for the bound (30). □

Condition (42) is generally difficult to check. A much simpler condition is expressed by the following Theorem.

Theorem 2. Decomposability (AWGN): The term $P(\underline{\hat{x}} \to \underline{\hat{z}})$ can be neglected in (30) without altering the inequality if there exist two sequences $\underline{z}'$ and $\underline{z}''$ different from $\underline{\hat{x}}$ and $\underline{\hat{z}}$, either in $\mathcal{X}^{\underline{S}}_{\underline{\hat{c}}}$ or in $\mathcal{X}^{\underline{S}}_{\underline{c}}$, such that

$$\begin{aligned} &1)\quad \mathrm{Re}\,[(\underline{z}' - \underline{\hat{x}})^\dagger(\underline{z}'' - \underline{\hat{x}})] \geq 0 \\ &2)\quad \underline{\hat{z}} = \underline{z}' + \underline{z}'' - \underline{\hat{x}}. \end{aligned} \tag{43}$$

Proof: Since (43) is translation-invariant, without loss of generality we can choose $\underline{\hat{x}} = \underline{0}$. Setting $\underline{y} \in \Gamma_{\underline{0},\underline{\hat{z}}}$, $\underline{y} \notin \Gamma_{\underline{0},\underline{z}'}$, and $\underline{y} \notin \Gamma_{\underline{0},\underline{z}''}$, we obtain

$$|\underline{y}|^2 \leq |\underline{y} - \underline{z}'|^2 \quad \text{and} \quad |\underline{y}|^2 \leq |\underline{y} - \underline{z}''|^2$$

or, equivalently,

$$2\,\mathrm{Re}(\underline{y}^\dagger \underline{z}') \leq |\underline{z}'|^2 \quad \text{and} \quad 2\,\mathrm{Re}(\underline{y}^\dagger \underline{z}'') \leq |\underline{z}''|^2.$$

Since $\underline{\hat{z}} = \underline{z}' + \underline{z}''$ and $\mathrm{Re}((\underline{z}')^\dagger \underline{z}'') \geq 0$, summing up the above inequalities we obtain

$$2\,\mathrm{Re}(\underline{y}^\dagger \underline{\hat{z}}) \leq |\underline{z}'|^2 + |\underline{z}''|^2 \leq |\underline{\hat{z}}|^2$$

contrary to the assumption that $\underline{y} \in \Gamma_{\underline{0},\underline{\hat{z}}}$. We conclude that $\underline{y}$ must be either in $\Gamma_{\underline{0},\underline{z}'}$ or $\Gamma_{\underline{0},\underline{z}''}$, and hence

$$\Gamma_{\underline{0},\underline{\hat{z}}} \subseteq \{\Gamma_{\underline{0},\underline{z}'} \cup \Gamma_{\underline{0},\underline{z}''}\}$$

The proof follows from Theorem 1. □

Remark: Theorem 2 is a generalization of the expurgation technique based on "decomposability" of error sequences, used by Verdú to bound the minimum BER of an uncoded multiuser system with antipodal binary modulation in AWGN [21].

Corollary 1. Extension to Fading Channels $(K < \infty)$*:* The term $P(\underline{\hat{x}} \to \underline{\hat{z}})$ can be neglected in (30) without changing the inequality if there exist two sequences $\underline{z}'$ and $\underline{z}''$, different from $\underline{\hat{x}}$ and $\underline{\hat{z}}$, either in $\mathcal{X}^{\underline{S}}_{\underline{\hat{c}}}$ or in $\mathcal{X}^{\underline{S}}_{\underline{c}}$, such that, for all components $\hat{x}_k$, $\hat{z}_k$, z'_k, and z''_k

$$\begin{aligned} &1)\quad \mathrm{Re}\,[(z'_k - \hat{x}_k)^\dagger(z''_k - \hat{x}_k)] \geq 0 \\ &2)\quad \hat{z}_k = z'_k + z''_k - \hat{x}_k. \end{aligned} \tag{44}$$

Proof: Condition (44) is the componentwise version of (43). Let g_k denote the (scalar) fading gain affecting the kth component of the received signal. If (44) holds, we get trivially that

$$\begin{aligned} &1)\quad \mathrm{Re}\,[(g_k z'_k - g_k \hat{x}_k)^\dagger(g_k z''_k - g_k \hat{x}_k)] \geq 0 \\ &2)\quad g_k \hat{z}_k = g_k z'_k + g_k z''_k - g_k \hat{x}_k. \end{aligned} \tag{45}$$

Then, let $\boldsymbol{G} = \mathrm{diag}\,(g_k \boldsymbol{I}_N)$, and consider the new signal sequences $\boldsymbol{G}\underline{\hat{x}}$, $\boldsymbol{G}\underline{\hat{z}}$, $\boldsymbol{G}\underline{z}'$, and $\boldsymbol{G}\underline{z}''$. We can extend (45) to sequences, which yields

$$\begin{aligned} &1)\quad \mathrm{Re}\,[(\boldsymbol{G}\underline{z}' - \boldsymbol{G}\underline{\hat{x}})^\dagger(\boldsymbol{G}\underline{z}'' - \boldsymbol{G}\underline{\hat{x}})] \geq 0 \\ &2)\quad \boldsymbol{G}\underline{\hat{z}} = \boldsymbol{G}\underline{z}' + \boldsymbol{G}\underline{z}'' - \boldsymbol{G}\underline{\hat{x}}. \end{aligned} \tag{46}$$

Hence, Theorem 2 can be applied for any fixed $\boldsymbol{G}$. Finally, we can eliminate the event $\{\underline{\hat{x}} \to \underline{\hat{z}}\}$ from the bound by: i) conditioning with respect to $\boldsymbol{G}$; ii) applying Theorem 2 and eliminating the region $\Gamma_{\boldsymbol{G}\underline{\hat{x}},\boldsymbol{G}\underline{\hat{z}}}$ for $\boldsymbol{G}$ given; iii) averaging the expurgated bound with respect to $\boldsymbol{G}$. □

1) Application to QAM Signal Sets with Gray Labeling: Application of Corollary 1 (and of Theorem 2) is particularly simple when $\mathcal{X}$ is a square M-QAM signal set with Gray labeling (see, for example, the 16QAM signal set represented in Fig. 2(b)). It is easy to verify that, with Gray labeling, each $\boldsymbol{x} \in \mathcal{X}^i_b$ has exactly a single nearest neighbor $\hat{\boldsymbol{z}} \in \mathcal{X}^i_{\bar{b}}$ (not necessarily at distance $d_{\min}$), for all $i = 1, \cdots, m$ and $b = 0, 1$. Moreover, for each $\boldsymbol{x} \in \mathcal{X}^i_b$, all the pairs $(\boldsymbol{x}, \boldsymbol{z})$ with $\boldsymbol{z} \in \mathcal{X}^i_{\bar{b}}$ and $\boldsymbol{z} \neq \hat{\boldsymbol{z}}$, satisfy condition (44) of Corollary 1. For example, if $\boldsymbol{x} \in \mathcal{X}^1_1$ corresponds to the point labeled 1011 in Fig. 2(b), $\hat{\boldsymbol{z}} \in \mathcal{X}^1_0$ to the point labeled 0011, and $\boldsymbol{z} \in \mathcal{X}^1_0$ to the point labeled 0010, then $\boldsymbol{z}'$ and $\boldsymbol{z}''$ correspond to the points labeled 0011 and 1010, respectively. For a given integer d, $\mathcal{X}^d$ is a hypercube carved from a cubic lattice of dimension d. Hence, condition (44) can be immediately extended to sequences in $\mathcal{X}^d$. We conclude that for all $\underline{\boldsymbol{x}} \in \mathcal{X}^{\underline{S}}_{\underline{c}}$ there exists only one relevant error event $\{\underline{\boldsymbol{x}} \to \underline{\hat{\boldsymbol{z}}}\}$, where $\underline{\hat{\boldsymbol{z}}}$ is the unique nearest neighbor of $\underline{\boldsymbol{x}}$ in $\mathcal{X}^{\underline{S}}_{\underline{\hat{c}}}$. By expurgating the BICM UB (32) we get the new bound

$$\begin{aligned} f(d, \mu, \mathcal{X}) &\leq m^{-d} \sum_{\underline{S}} 2^{-d} \sum_{\underline{U}} 2^{-d(m-1)} \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}^{\underline{S}}_{\underline{c}}} P(\underline{\boldsymbol{x}} \to \underline{\hat{\boldsymbol{z}}}) \\ &\triangleq f_{\mathrm{ex}}(d, \mu, \mathcal{X}). \end{aligned} \tag{47}$$

This will be referred to as *BICM Expurgated Bound* (EX). In order to obtain a computationally efficient form of it, we use once again a Laplace-transform approach. Let $\Delta(\boldsymbol{x}, \hat{\boldsymbol{z}})$ denote the metric difference

$$\Delta(\boldsymbol{x}, \hat{\boldsymbol{z}}) = \log p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) - \log p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \hat{\boldsymbol{z}})$$

between the branch metric of $\boldsymbol{x}$ and the branch metric of its (unique) nearest neighbor $\hat{\boldsymbol{z}} \in \mathcal{X}^i_{\bar{b}}$. By using (34) and exchanging the sums and the integral, we get

$$f_{\mathrm{ex}}(d, \mu, \mathcal{X}) = \frac{1}{2\pi j} \int_{\alpha - j\infty}^{\alpha + j\infty} \psi_{\mathrm{ex}}(s)^d \, \frac{ds}{s} \tag{48}$$

where we have defined the function $\psi_{\mathrm{ex}}(s)$ as

$$\psi_{\mathrm{ex}}(s) \triangleq \frac{1}{m 2^m} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{\boldsymbol{x} \in \mathcal{X}^i_b} \Phi_{\Delta(\boldsymbol{x}, \hat{\boldsymbol{z}})}(s). \tag{49}$$

The BICM EX is formally similar to the BICM UB, but it includes only one error event $\{\underline{\boldsymbol{x}} \to \underline{\hat{\boldsymbol{z}}}\}$ for each possible transmitted sequence $\underline{\boldsymbol{x}}$, rather than all the possible $2^{d(m-1)}$ error events $\{\underline{\boldsymbol{x}} \to \underline{\boldsymbol{z}}\}$, for all $\underline{\boldsymbol{z}} \in \mathcal{X}^{\underline{S}}_{\underline{\hat{c}}}$. Computational experience has shown that (48), when used with other signal sets, often provides a very tight approximation to the PEP. Actually, as we prove in the following subsection, with Gray labeling and AWGN the approximation obtained from BICM is asymptotically tight. Numerical evaluation of (48) can be efficiently performed by using the Gauss–Chebyshev method of [4].

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

Example 3 (Continued): If we apply the BICM EX to the case of uncoded 4PSK with Gray labeling over AWGN we obtain

$$P_{\text{bin}} \leq f_{\text{ex}}(1, \mu, \mathcal{X}) = Q(1/\sigma)$$

which is tight (see Example 3 of Section IV-B).

2) Asymptotic Behavior of the PEP: Here we establish the behavior of $f(d, \mu, \mathcal{X})$ as $\sigma \to 0$ with fading channels with perfect CSI. This provides some insight into the asymptotic performance of BICM for high SNR. For Rician fading channels with perfect CSI we get [20]

$$\Phi_{\Delta(\boldsymbol{x},\boldsymbol{z})}(s) = \frac{\exp\left(-\frac{s(1-s\sigma^2)K|\boldsymbol{x}-\boldsymbol{z}|^2/(K+1)}{1+s(1-s\sigma^2)|\boldsymbol{x}-\boldsymbol{z}|^2/(K+1)}\right)}{1+s(1-s\sigma^2)|\boldsymbol{x}-\boldsymbol{z}|^2/(K+1)}. \tag{50}$$

For AWGN $f_{\text{ub}}(d, \mu, \mathcal{X})$ and $f_{\text{ex}}(d, \mu, \mathcal{X})$ can be written as

$$\begin{aligned} f_{\text{ub}}(d, \mu, \mathcal{X}) &= m^{-d} \sum_{\underline{\boldsymbol{S}}} 2^{-d} \sum_{\underline{\boldsymbol{U}}} 2^{-d(m-1)} \\ &\quad \times \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \sum_{\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}} Q\left(\frac{|\underline{\boldsymbol{x}} - \underline{\boldsymbol{z}}|}{\sqrt{2}\sigma}\right) \end{aligned} \tag{51}$$

$$\begin{aligned} f_{\text{ex}}(d, \mu, \mathcal{X}) &= m^{-d} \sum_{\underline{\boldsymbol{S}}} 2^{-d} \sum_{\underline{\boldsymbol{U}}} 2^{-d(m-1)} \\ &\quad \times \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} Q\left(\frac{|\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{z}}}|}{\sqrt{2}\sigma}\right). \end{aligned} \tag{52}$$

The minimum possible Euclidean distance between any $\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$ and $\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$, for all possible sequences of positions $\underline{\boldsymbol{S}}$, is $\sqrt{d}d_{\min}$, where $d_{\min}$ denotes the minimum Euclidean distance of $\mathcal{X}$. From (51) we get

$$\begin{aligned} &\lim_{\sigma \to 0} \frac{f_{\text{ub}}(d, \mu, \mathcal{X})}{Q(\sqrt{d}d_{\min}/(\sqrt{2}\sigma))} \\ &\quad = m^{-d} \sum_{\underline{\boldsymbol{S}}} 2^{-d} \sum_{\underline{\boldsymbol{U}}} 2^{-d(m-1)} \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} N_{\min}(d, \underline{\boldsymbol{x}}) \\ &\quad = N_{\min}(d) \end{aligned} \tag{53}$$

where $N_{\min}(d, \underline{\boldsymbol{x}})$ is the number of neighbors $\underline{\boldsymbol{z}} \in \mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$ at distance $\sqrt{d}d_{\min}$ from $\underline{\boldsymbol{x}}$ and $N_{\min}(d)$ is the expected number of nearest neighbors of an error event of Hamming distance d, averaged over $\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$, $\underline{\boldsymbol{S}}$, and $\underline{\boldsymbol{U}}$. In order to compute $N_{\min}(d)$, we write

$$N_{\min}(d, \underline{\boldsymbol{x}}) = \prod_{k=1}^{d} N_{\min}(1, \boldsymbol{x}_k).$$

Hence, from the right-hand side of (51) we get

$$N_{\min}(d) = [N_{\min}(1)]^d \tag{54}$$

where

$$N_{\min}(1) \triangleq \frac{1}{m2^m} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{\boldsymbol{x} \in \mathcal{X}_b^i} N_{\min}(1, \boldsymbol{x}) \tag{55}$$

is the average number of signals at distance $d_{\min}$ in the complement subset of a signal $\boldsymbol{x}$. Note that $N_{\min}(1)$ depends only on $\mathcal{X}$ and on the labeling μ. From (54) we may observe the impact of the value of $N_{\min}(1)$ on the performance of BICM: in fact, if this is greater than 1 then $N_{\min}(d)$ increases exponentially with the Hamming distance d, thus impairing the performance of the coding scheme at moderate SNR. If, on the contrary, $N_{\min}(1) < 1$, increasing d does decrease the value of $N_{\min}(d)$, and hence improves the performance. This fact is further discussed in the following, where we show how the choice of the labeling map μ affects the value of $N_{\min}(1)$.

Let now $\mathcal{D}$ denote the event that a signal sequence $\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$ has a nearest neighbor at distance $\sqrt{d}d_{\min}$ in $\mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$, and let $\mathcal{I}_{\mathcal{D}}(\underline{\boldsymbol{x}})$ denote the indicator function of $\mathcal{D}$. Then, from (51) we get

$$\begin{aligned} &\lim_{\sigma \to 0} \frac{f_{\text{ex}}(d, \mu, \mathcal{X})}{Q(\sqrt{d}d_{\min}/(\sqrt{2}\sigma))} \\ &\quad = m^{-d} \sum_{\underline{\boldsymbol{S}}} 2^{-d} \sum_{\underline{\boldsymbol{U}}} 2^{-d(m-1)} \sum_{\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}} \mathcal{I}_{\mathcal{D}}(\underline{\boldsymbol{x}}) \\ &\quad = P(\mathcal{D}) \leq 1. \end{aligned} \tag{56}$$

In the case of Gray labeling we have

Theorem 3. Asymptotically Tight Bound and Approximation (AWGN): If μ is a Gray labeling for the signal set $\mathcal{X}$, then $N_{\min}(d) \leq 1$, and

$$\lim_{\sigma \to 0} \frac{f_{\text{ub}}(d, \mu, \mathcal{X})}{f(d, \mu, \mathcal{X})} = \lim_{\sigma \to 0} \frac{f_{\text{ex}}(d, \mu, \mathcal{X})}{f(d, \mu, \mathcal{X})} = 1. \tag{57}$$

Proof: If μ is a Gray labeling for $\mathcal{X}$, by definition $N_{\min}(d, \underline{\boldsymbol{x}})$ is either 1 or 0, hence $N_{\min}(d, \underline{\boldsymbol{x}}) = \mathcal{I}_{\mathcal{D}}(\underline{\boldsymbol{x}})$ and

$$N_{\min}(d) = P(\mathcal{D}) \leq 1. \tag{58}$$

In order to prove (57), we need a lower bound on $f(d, \mu, \mathcal{X})$. We follow the technique of [21] and consider a *genie* decoder with additional side information. In particular, suppose that, for each transmitted signal sequence $\underline{\boldsymbol{x}} \in \mathcal{X}_{\underline{\boldsymbol{c}}}^{\underline{\boldsymbol{S}}}$ the genie decoder knows that the only possible concurrent signal sequence is $\hat{\underline{\boldsymbol{z}}}$, i.e., the unique sequence at minimum distance from $\underline{\boldsymbol{x}}$ in the complement subset $\mathcal{X}_{\hat{\underline{\boldsymbol{c}}}}^{\underline{\boldsymbol{S}}}$. If $|\hat{\underline{\boldsymbol{z}}} - \underline{\boldsymbol{x}}| > \sqrt{d}d_{\min}$, the PEP of the genie decoder is trivially bounded from below by zero. Hence, we can write

$$f(d, \mu, \mathcal{X}) \geq P(\mathcal{D}) Q\left(\frac{\sqrt{d}d_{\min}}{\sqrt{2}\sigma}\right). \tag{59}$$

If $\mathcal{X}$ is a squared QAM signal set, we can write

$$\begin{aligned} P(\mathcal{D}) Q\left(\frac{\sqrt{d}d_{\min}}{\sqrt{2}\sigma}\right) &\leq f(d, \mu, \mathcal{X}) \leq f_{\text{ex}}(d, \mu, \mathcal{X}) \\ &\leq f_{\text{ub}}(d, \mu, \mathcal{X}). \end{aligned} \tag{60}$$

By taking the limit for $\sigma \to 0$ and by using (53), (56), and (58) we obtain (57). If $\mathcal{X}$ is not a squared QAM signal set, the order of the middle inequality in (60) may be reversed (i.e., it may be $f_{\text{ex}}(d, \mu, \mathcal{X}) \leq f(d, \mu, \mathcal{X})$). However, it is immediate to check that the left and the right inequalities still hold, thus giving (57) as $\sigma \to 0$. □

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

For any other labeling, the asymptotic ratio between the PEP upper and lower bounds is $N_{\min}(d)/P(\mathcal{D}) > 1$. Hence, Theorem 3 shows a sort of asymptotic optimality of Gray labeling for BICM.

For $K < \infty$, the asymptotic analysis is more complicated, because a closed-form expression for the PEP $P(\underline{\boldsymbol{x}} \to \underline{\hat{\boldsymbol{z}}})$ for arbitrary $\underline{\boldsymbol{x}}$, $\underline{\hat{\boldsymbol{z}}}$, and K is missing, and because, if $K < \infty$, $P(\underline{\boldsymbol{x}} \to \underline{\hat{\boldsymbol{z}}})$ does not decrease exponentially with SNR. For these reasons, we restrict our analysis to the BICM EX and we use the Chernoff bound (see (38)). From (38) and (50) we obtain

$$\min_{0<\alpha<\alpha_2} \Phi_{\Delta(\underline{\boldsymbol{x}},\underline{\hat{\boldsymbol{z}}})}(\alpha) = \prod_{k=1}^{d} \frac{\exp\left(-\frac{K|\boldsymbol{x}_k-\hat{\boldsymbol{z}}_k|^2/(4\sigma^2(K+1))}{1+|\boldsymbol{x}_k-\hat{\boldsymbol{z}}_k|^2/(4\sigma^2(K+1))}\right)}{1+|\boldsymbol{x}_k-\hat{\boldsymbol{z}}_k|^2/(4\sigma^2(K+1))}. \tag{61}$$

By assuming that $\mathcal{X}$ is a square QAM constellation with Gray labeling, by substituting (61) into the BICM EX (47), and by choosing σ so small that $1 \ll |\boldsymbol{x}_k - \hat{\boldsymbol{z}}_k|^2/(4\sigma^2(K+1))$ for all $k = 1, \cdots, d$, we get

$$f_{\text{ex}}(d, \mu, \mathcal{X}) \sim \left(\frac{4(K+1)e^{-K}}{d_{\text{h}}^2/\sigma^2}\right)^d \tag{62}$$

where

$$d_{\text{h}}^2 = \left(\frac{1}{m2^m}\sum_{i=1}^{m}\sum_{b=0}^{1}\sum_{\boldsymbol{x}\in\mathcal{X}_b^i}\frac{1}{|\boldsymbol{x}-\hat{\boldsymbol{z}}|^2}\right)^{-1}. \tag{63}$$

Hence, for any given $K < \infty$, as $\sigma \to 0$ the function $f_{\text{ex}}(d, \mu, \mathcal{X})$ depends on the inverse dth power of SNR. In the fading case there is no single dominant term, and therefore the union bound may not be tight as $\sigma \to 0$. Results of Section V show that the BICM EX is very close to the real PEP (obtained by simulation) also in the fading case, while the BICM UB is extremely loose at all SNR's.

The coefficient d_{h}^2 can be interpreted as the *harmonic mean* of the minimum squared Euclidean distance between complementary subsets of $\mathcal{X}$. By comparing the left-hand of (62) with the Chernoff bound on the PEP of a CM scheme in Rician fading (see, for example, [14, p. 162]), we observe that $(d_{\text{h}}^2)^d$ plays the role of the so-called *squared product distance* of the error event.

The role of the two basic parameters $N_{\min}(1)$ and d_{h}^2 in the design of BICM will be emphasized in next section.

V. Design Guidelines

The analysis of previous sections shows that the performance of BICM with ideal interleaving depends mainly on the following parameters. i) The minimum binary Hamming distance of $\mathcal{C}$, denoted in the following by $d_2(\mathcal{C})$. ii) The minimum Euclidean distance $d_{\min}$ of $\mathcal{X}$. iii) The average number of nearest neighbors $N_{\min}(1)$, defined in (55). iv) The harmonic mean of the minimum squared distance d_{h}^2, defined in (63).

On the AWGN channel with coherent detection (perfect CSI), the performance of BICM is asymptotically dominated by the error event at minimum Euclidean distance

$$d_{\text{E}} = \sqrt{d_2(\mathcal{C})}\, d_{\min}. \tag{64}$$

From the union bound (26), the error coefficient corresponding to the distance d_{E} is $W_I(d_2(\mathcal{C}))[N_{\min}(1)]^{d_2(\mathcal{C})}$ divided by k_c or by K_c, in the case of convolutional or block codes, respectively.

In the case of independent (or ideally interleaved) Rayleigh fading channel with perfect CSI, the asymptotic error performance of any code is proportional to the inverse of SNR raised to a power equal to the *code diversity*, i.e., the minimum Hamming distance between signal sequences produced by the encoder/modulator [14]. We showed that in BICM this code diversity equals $d_2(\mathcal{C})$. Recalling the definition of the spectral efficiency R, with PSK/QAM and Gray labeling, and letting $\sigma \to 0$ and $\text{SNR} = 1/\sigma^2 = RE_b/N_0$, from (62) we can write

$$\log_{10} P_{\text{b}} \simeq -d_2(\mathcal{C})\left[\left(Rd_{\text{h}}^2\right)_{\text{dB}} + \left(\frac{E_b}{N_0}\right)_{\text{dB}}\right]\Big/10 + \text{const.} \tag{65}$$

Thus for high values of E_b/N_0, the BER curve on a logarithmic scale approaches a straight line with slope proportional to $-d_2(\mathcal{C})$, translated horizontally by the offset $(Rd_{\text{h}}^2)_{\text{dB}}$. As we shall see in an example, this rule of thumb turns out to be useful for a quick comparison between different BICM schemes over the Rayleigh fading channel.

Consider next the case where $\mathcal{C}$ is a convolutional code. The decoder complexity can be measured by the *trellis complexity* $\tau = N_S N_B$, where $N_S = 2^v$ denotes the number of states in the minimal encoder for $\mathcal{C}$ (v is the encoder memory), and $N_B = 2^{k_c}$ is the number of branches merging into each state at each trellis step (k_c is the number of encoder inputs). The complexity of punctured convolutional codes is the same τ of the low-rate code; they are derived from [7].

The design of BICM can be summarized as follows:

1) Given a signal set $\mathcal{X}$ of cardinality $M = 2^m$ and dimension N, a desired decoder complexity τ, and a required spectral efficiency R, choose the best binary convolutional code of rate $r_c = k_c/n_c$ such that $R = r_c m/N$, with complexity $\leq \tau$.
2) Depending on the application (AWGN, Rayleigh fading, etc.) and on the detection technique (perfect CSI, no CSI, etc.), choose the labeling map μ according to some optimality criterion. We showed that, for perfect CSI, practical and sensible criteria are the minimization of $N_{\min}(1)$ and the maximization of d_{h}^2. Table I lists $N_{\min}(1)$ and d_{h}^2 for some PSK and QAM signal sets and some labelings (observe that Gray and quasi-Gray labelings provide values of $N_{\min}(1)$ never exceeding 1, unlike other labelings).

In this way, the binary code is picked independently of μ and $\mathcal{X}$. Since the best convolutional codes are known for many rates and complexities (see [17], [11], [7], and references therein) normally no search for new binary codes is needed. We mention in passing that BICM easily allows the concatenation of any given signal set with the powerful

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

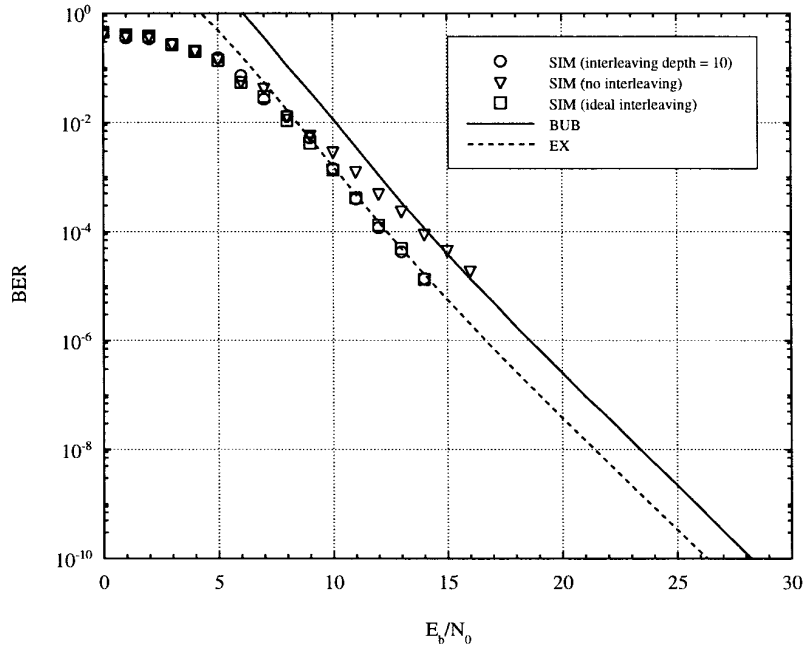

Fig. 10. BER of BICM 8PSK code obtained from the optimal 8-state, rate-$2/3$ code and Gray labeling. Rayleigh fading with perfect CSI.

TABLE I
VALUES OF $N_{\min}(1)$ AND d_{h}^2 FOR SOME PSK AND QAM SIGNAL SETS. (THE SIGNAL SET AVERAGE ENERGY IS NORMALIZED TO 1. LABELS "$\sim G$" AND "W" DENOTE QUASI-G RAY AND THE LABELING OF 32QAM PROPOSED BY WEI IN [25] FOR DESIGNING UNEQUAL ERROR PROTECTION CM)

$\mathcal{X}$	μ	$N_{\min}(1)$	d_{h}^2
4PSK	Gray	1	2
	SP	1.5	2
8PSK	Gray	0.666	0.7664
	SP	1.166	0.664
8QAM	$\sim$ G	0.9167	0.96
	SP	1.333	0.96
16QAM	Gray	0.75	0.4923
	SP	1.75	0.4414
32QAM	$\sim$ G	0.75	0.2788
	W	1.2	0.2133
64QAM	Gray	0.5833	0.1442
128QAM	$\sim$ G	0.6339	0.077
256QAM	Gray	0.4688	0.043

class of binary parallel concatenated codes [15] (also known as "turbo codes").

A. Numerical Results

In this subsection we show a selection of numerical results aimed at illustrating some features and applications of BICM. In the figures, curves marked by BUB denote union-Bhattacharyya bound, UB the BICM union bound, EX the BICM expurgated bound (or approximation), and SIM computer simulation. All the simulation results presented hereafter were obtained by using the suboptimal branch metric (9).

1) Effect of Finite-Depth Interleaving: Here we prove that interleaving is indeed necessary, although it need not be very deep if the channel has a short memory. Fig. 10 shows the BER of BICM over independent Rayleigh fading with perfect CSI, where $\mathcal{X}$ is 8PSK, μ is Gray labeling, and $\mathcal{C}$ is the optimal 8-state, rate-$2/3$ code used by Zehavi [26]. The only difference between this and Zehavi's BICM scheme is that here we use a single bit interleaver instead of separate interleavers for the three encoder outputs. Simulation in the case of ideal interleaving shows excellent agreement with the BICM EX approximation. The Bhattacharyya union bound is about 2 dB away from the true BER, as typical of fading channels. Zehavi's analysis based on the Chernoff bound [26] shows about the same gap from simulation. Hence, the usefulness of the tighter bounds developed in this paper is apparent. Simulations are also shown in the case of no interleaving and with interleaving depth equal to 10. Note that, since the Rayleigh fading channel used here is memoryless, the only effect of interleaving is to break the correlation introduced by the modulation, which carries three bits in a single transmitted signal. Absence of interleaving degrades the BER: however, a relatively short interleaver is sufficient to approach the ultimate performance.

2) Gray Labeling: Here we prove that SP labeling is poor for BICM. Fig. 11 shows the BER of BICM over AWGN with perfect CSI, where $\mathcal{X}$ is 16QAM and $\mathcal{C}$ is the *de facto* standard 64-state, rate-$1/2$ binary convolutional code with (octal) generators $(133, 171)$ [17, pp. 466–471]. Gray and SP labelings are compared. Note that BICM EX is a true

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

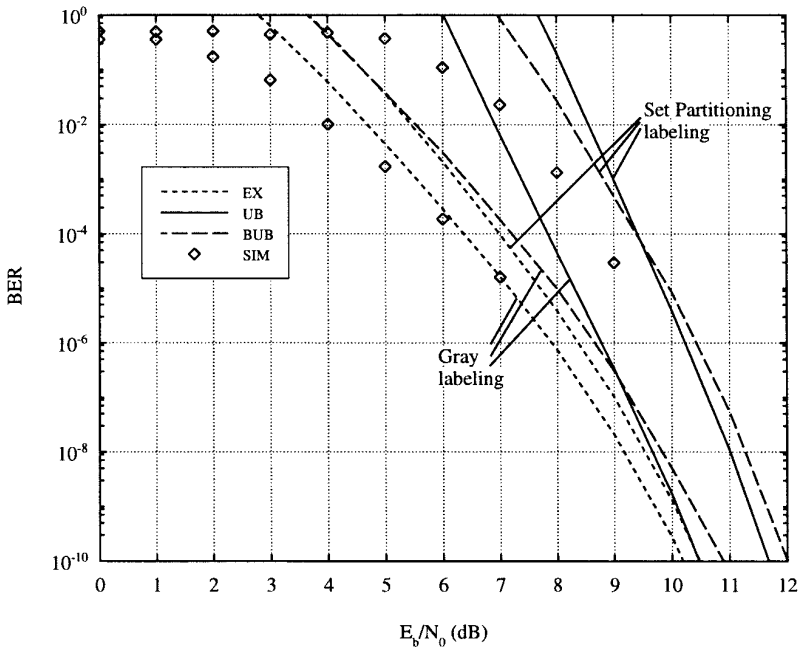

Fig. 11. BER of BICM 16QAM code obtained from the optimal 64-states rate-1/2 code. AWGN channel.

(very tight) bound only for Gray labeling. With SP labeling, the BICM EX curve underestimates the actual BER. This is because, without Gray labeling, there exist many more nearest neighbors for each transmitted signal sequence, while the BICM EX counts just one neighbor. Hence, if the labeling is not Gray, the BICM EX curve yields too optimistic values. However, the Bhattacharyya union bound and the BICM UB always provides an upper bound (also for the SP labeling), although the latter is rather loose in the range of BER values of practical interest.

3) BICM PSK/QAM Codes for the Fading Channel: Fig. 12 shows the BER over Rayleigh fading channels of BICM codes obtained by concatenating the same rate-$1/2$ code of the previous example with different signal sets. Gray labeling is used whenever possible. When $\mathcal{X}$ does not admit Gray labeling, a quasi-Gray labeling is chosen (i.e., a labeling that minimizes the number of points with more than one nearest neighbor whose label differ by more than one bit). Since the code is the same, all the BER curves have asymptotically the same slope. The E_b/N_0 (in decibels) gap between them can be evaluated from (65). The differences between the values of $(Rd_{\mathrm{h}}^2)_{\mathrm{dB}}$ and those pertaining to 4PSK are reported in Table II. By comparing these values with the curves in Fig. 12, we note how (65) gives a quite accurate prediction. Note also that BICM EX gives true upper bounds for Gray labeling, but only an approximation in the quasi-Gray case.

A method for increasing code diversity with BICM consists of concatenating an expanded signal set to a low-rate code [13]. For given R and τ, this leads to asymptotically steeper BER curves. However, signal set expansion reduces the value of d_{h}^2, so that this technique may not provide a better performance in the range of BER values of interest. The

TABLE II
SPECTRAL EFFICIENCIES AND DIFFERENCES IN THE VALUES OF $(Rd_{\mathrm{h}}^2)_{\mathrm{dB}}$ FOR THE BICM CODES OF FIG. 12

$\mathcal{X}$	R (bit/dim)	E_b/N_0 gap (dB)
4PSK	1	0
8PSK	1.5	2.4
16QAM	2	3.07
32QAM	2.5	4.57
64QAM	3	6.65
128QAM	3.5	8.7
256QAM	4	10.65

crossover point between different BICM codes over Rayleigh fading can be coarsely estimated by finding the intersection between the straight lines defined by (65) (by neglecting the "const." terms). Fig. 13 shows two examples of BICM design based on signal set expansion. We considered two BICM schemes with $R = 2$ bit/dim and $\tau = 128$, one obtained by concatenating to 8PSK the best rate-$2/3$, 64-state punctured code [7] and the other obtained by concatenating to 16QAM the best rate-$1/2$, 64-state code of previous example (both with Gray labeling). From (65) we estimate the crossover at $E_b/N_0 = 2.95$ dB, which indicates that in this case signal set expansion yields a coding gain at almost any BER of interest. Next, we considered two BICM codes with $R = 1$ bit/dim and $\tau = 128$, one obtained by concatenating to 4PSK the same best rate-$1/2$, 64-state code and the other obtained by concatenating to 16QAM the best rate-$1/4$, 64-state code [17, pp. 466–471] (both with Gray labeling). From (65) we estimate the crossover at $E_b/N_0 = 9.16$ dB, which indicates that in this case signal

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

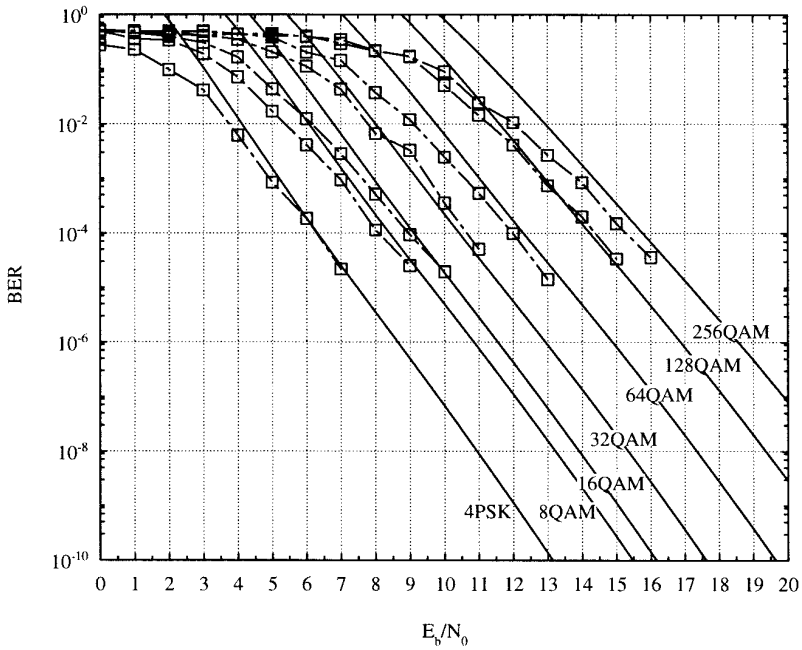

Fig. 12. BER of BICM obtained from the optimal 64-state rate-$1/2$ code and PSK/QAM signal sets. Rayleigh fading with perfect CSI.

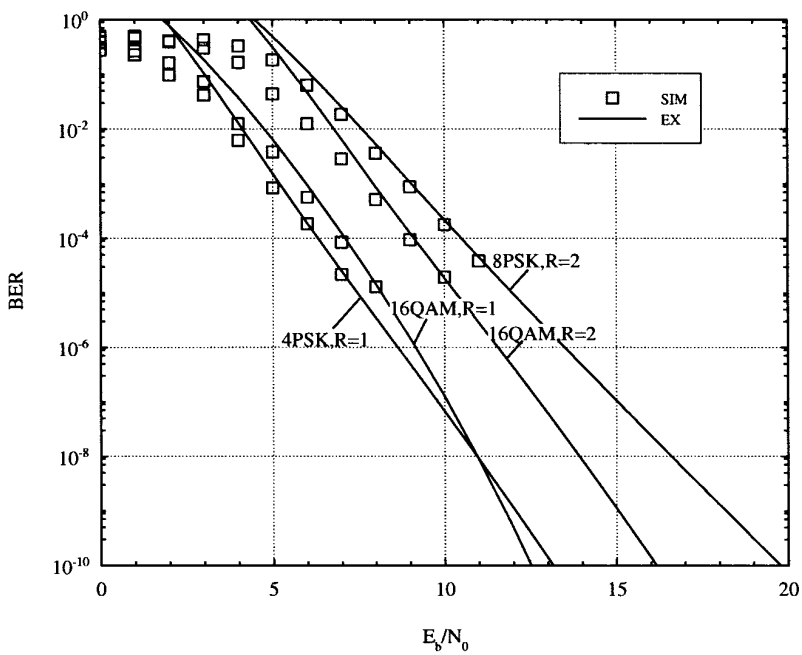

Fig. 13. BER of BICM for $R = 2$ and $R = 1$ with $\tau = 128$ (64 states). The codes are obtained by concatenating the best rate-$2/3$ punctured code with 8PSK, the best rate-$1/2$ code with 16QAM, the best rate-$1/2$ code with 4PSK and the best rate-$1/4$ code with 16QAM. Rayleigh fading with perfect CSI.

set expansion yields a coding gain only at very low BER values. The actual BER curves intersect at $E_b/N_0 = 11$ dB.

Finally, we can compare BICM and TCM by using bounds on the achievable minimum Euclidean distance and code diversity. For BICM, we can use Heller bound and its extension to rate k_c/n_c codes as provided in [11]. For TCM, sphere-packing bounds on d_{E} can be found in [5, Ch. 4] and references therein. In the case of Ungerboeck's TCM (i.e., TCM obtained from a binary encoder for $\mathcal{C}$), the code diversity is given by the M-ary Hamming distance $d_M(\mathcal{C})$ of $\mathcal{C}$ ($M = 2^m$), considered as a k_c-to-1 binary input M-ary output code. In this case, we can use the upper bound on the M-ary Hamming distance of [16]

$$d_M(\mathcal{C}) \leq \min\left\{\left\lfloor \frac{v}{k_c} \right\rfloor + 1, \min_{tk_c \geq m}\left\{\frac{M-1}{M}\frac{2^{tk_c}}{2^{tk_c}-1}\left(\left\lfloor \frac{v}{k_c} \right\rfloor + t\right)\right\}\right\} \tag{66}$$

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

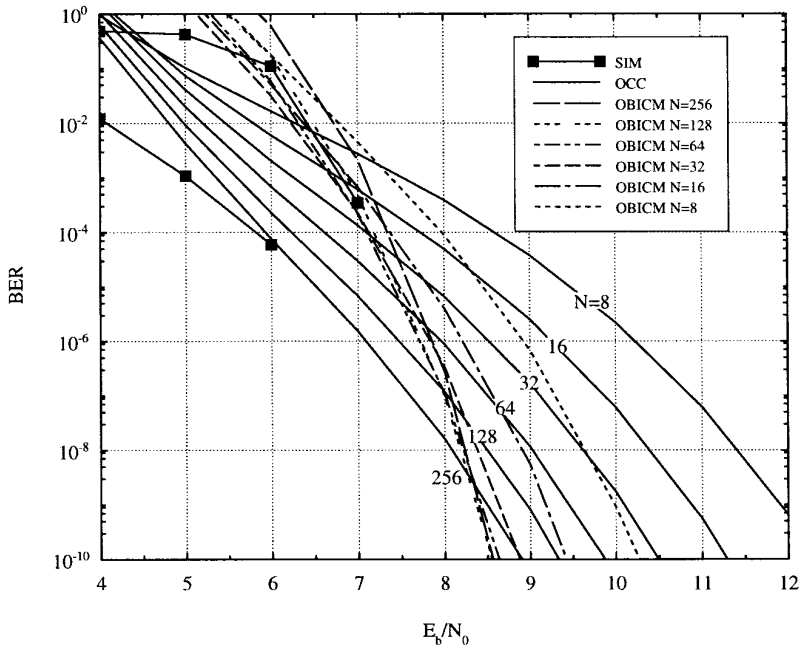

Fig. 14. OBICM and OCC over AWGN with noncoherent detection. Simulation results are shown for $N = 256$ only.

TABLE III
UPPER BOUNDS TO MINIMUM EUCLIDEAN DISTANCE AND CODE DIVERSITY FOR TCM AND BICM CODES FOR 16QAM (AVERAGE ENERGY NORMALIZED TO 1) WITH $R = 3$ bit/dim

Encoder	BICM		TCM	
memory	d_E^2	$d_2(\mathcal{C})$	d_E^2	$d_M(\mathcal{C})$
2	1.2	3	2	1
3	1.6	4	2.4	2
4	1.6	4	2.8	2
5	2.4	6	3.2	2
6	2.4	6	3.6	3
7	3.2	8	3.6	3
8	3.2	8	4	3

where v is the encoder memory (i.e., the number of states is $N_S = 2^v$). In particular, the code diversity of TCM is bounded from above by $\lfloor v/k_c \rfloor + 1$, irrespective of the encoder rate k_c/m. Table III shows the bounds on the minimum Euclidean distance and on the minimum code diversity of BICM and TCM, for 16QAM codes with $R = 3$ bit/dim. As expected from the discussions above, for the same complexity BICM achieves better code diversity, while TCM achieves better minimum Euclidean distance.

BICM Codes for DS/SSMA with Noncoherent Detection: Viterbi's orthogonal convolutional codes (OCC) have been proposed for coding and spreading in DS/SSMA. These codes can be seen as TCM schemes, where a binary encoder of rate $1/m$ and 2^{m-1} states is concatenated to a Hadamard modulator which generates an $N = 2^m$-dimensional orthogonal signal set. The binary encoder is a simple shift register, and its generator matrix (expressed in polynomial form) is

$$G(D) = [1, D, D^2, \cdots, D^{m-1}].$$

The decoder complexity is $\tau = 2^m = N$. In a practical implementation, the branch metrics can be computed by a fast Hadamard transform, applied to the received vector $\boldsymbol{y}$, the N-sample sequence output by a chip matched filter [22].

With noncoherent detection and no CSI, the branch metrics can be approximated by taking the squared magnitude of the outputs of the Hadamard transform [17], [12], similarly to the usual squared envelope detector used for orthogonal noncoherent FSK [8]. From (66), we get immediately that OCC's have maximal M-ary free Hamming distance $d_M(\mathcal{C}) = m$ (where, in this case, $M = 2^m = N$). Hence, the code diversity is maximized by OCC among all the TCM schemes with given complexity $\tau = 2^m$. The path at M-ary distance m from the all-zero path is originated by a single "1" entering the shift register. Hence, OCCs have $d_2(\mathcal{C}) = d_M(\mathcal{C})$, so that they are not suited to BICM. In order to design a BICM scheme for orthogonal signal sets (OBICM) with the same decoding complexity $\tau = 2^m$ of OCC's (and the same demodulator described above), we pick the optimum binary convolutional codes of rate $1/m$ and 2^{m-1} states [17, pp. 466–471]. After ideal bit interleaving, m-bit labels are mapped onto the N-ary orthogonal signal set. The resulting spectral efficiency is still, as for OCC, $R = 1/N$ bit/dim, but now the code diversity is $d_2(\mathcal{C})$ (see Table IV). Figs. 14 and 15 show the BER of OCC and OBICM with noncoherent detection over AWGN and Rayleigh fading, respectively, for $N = 8, 16, 32, 64, 128, 256$. In the case of OCC, we can evaluate the union bound with exact PEP computations obtained by following [4] (these curves are labeled "TUB"). As for OBICM, we used our Bhattacharyya union bound, which is fairly tight in this case.

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

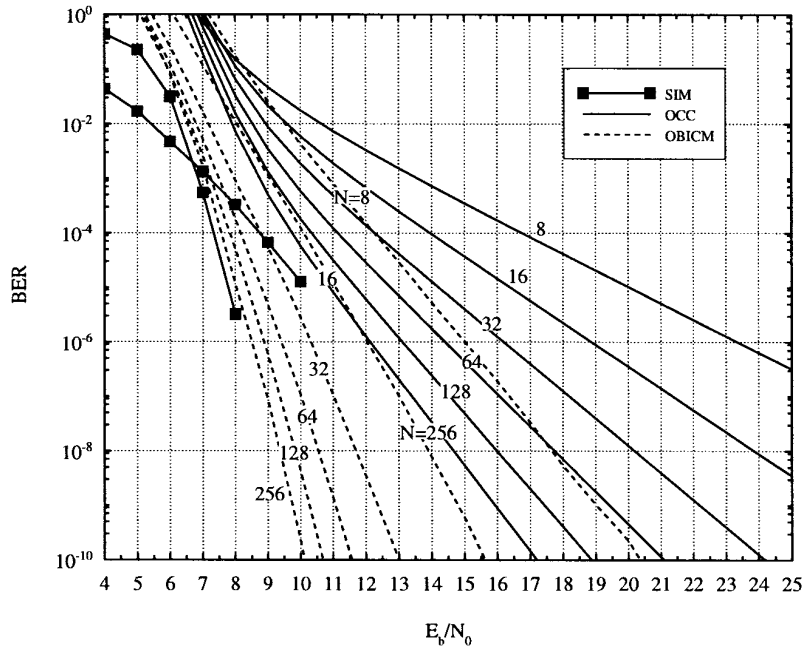

Fig. 15. OBICM and OCC over Rayleigh fading with noncoherent detection and no CSI. Simulation results are shown for $N = 256$ only.

TABLE IV
CODE DIVERSITY OF OBICM AND OCC FOR THE SAME DECODER COMPLEXITY AND RATE.

N	N_S	OBICM	OCC
8	4	8	3
16	8	13	4
32	16	20	5
64	32	27	6
128	64	36	7
256	128	45	8

OBICM compares favorably with OCC only for low N over AWGN. As N increases, OCC perform better, especially for BER $\geq 10^{-5}$. On the contrary, over Rayleigh fading, OBICM performs better than OCC. Note also that, due to its high code diversity, low-rate OBICM yields almost the same performance over AWGN and Rayleigh fading. For example, for $N = 256$ the performance loss at BER $= 10^{-5}$ due to Rayleigh fading with respect to AWGN is only 0.5 dB, while the corresponding loss of OCC with $N = 256$ is about 4.5 dB. This makes OBICM an interesting solution for coded DS/SSMA with noncoherent detection over channels where fading may range from AWGN to Rayleigh depending on the propagation environment.

VI. CONCLUSION

The main theme of this paper is that on some channels the separation of demodulation and decoding might be beneficial, provided that the encoder output is interleaved bit-wise and a suitable soft-decision metric is used in the Viterbi decoder. A comprehensive analysis of BICM, based on channel capacity and cutoff rate, shows this in information-theoretical terms. Optimum and simpler, suboptimum bit metrics are derived for channels with and without state information at the receiver. The central role of the labeling map is pinpointed, while an extensive error probability analyisis, which includes the derivation of sundry bounds and approximations, leads to design guidelines for BICM schemes. A comprehensive set of results suggests an array of possible applications.

REFERENCES

[1] S. A. Al-Semari and T. Fuja, "Bit interleaved I-Q TCM," in *ISITA'96* (Victoria, B.C., Sept. 17–20, 1996).

[2] A. Aoyama, T. Yamazato, M. Katayama, and A. Ogawa, "Performance of 16-QAM with increased diversity on Rayleigh fading channels," in *Proc. Int. Symp. Information Theory and Its Applications* (Sydney, Australia, Nov. 20–24, 1994), pp. 1133–1137.

[3] E. Biglieri, G. Caire, and G. Taricco, "Error probability over fading channels: A unified approach," to be published in *Europ. Trans. Commun.*, July 1997.

[4] E. Biglieri, G. Caire, G. Taricco, and J. Ventura, "Simple method for evaluating error probabilities," *Electron. Lett.*, vol. 32, no. 3, pp. 191–192, Feb. 1996.

[5] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. New York: MacMillan, 1991.

[6] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 502–518, Mar. 1996.

[7] J. B. Cain, G. C. Clark, and J. M. Geist, "Punctured convolutional codes of rate $(n-1)/n$ and simplified maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 97–100, Jan. 1979.

[8] G. Caire, J. Ventura, and E. Biglieri, "Coded and pragmatic-coded orthogonal modulation for the fading channel with non-coherent detection," in *Proc. IEEE Int. Conf. Communications (ICC'95)* (Seattle, WA, June 18–22), 1995.

[9] G. Caire, G. Taricco, and E. Biglieri, "Capacity and cut-off rate of bit-interleaved channels," in *Proc. Int. Symp. Information Theory and Its Applications ISITA'96* (Victoria, BC, Canada, Sept. 17–20, 1996).

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] D. G. Daut, J. W. Modestino, and L. D. Wismer, "New short constraint length convolutional code constructions for selected rational rates," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 794–800, Sept. 1982.

[12] D. Divsalar and M. Simon, "Maximum-likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channels—Metrics and performance," *IEEE Trans Commun.*, vol. 42, Jan. 1994.

[13] U. Hansson and T. Aulin, "Channel symbol expansion diversity—Improved coded modulation for the Rayleigh fading channel," presented at the Int. Conf. Communications, ICC'96, Dallas TX, June 23–27, 1996.

[14] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. New York: Kluwer, 1994.

[15] S. Le Goff, A. Glavieux, and C. Berrou, "Turbo-codes and high spectral efficiency modulation," in *Int. Conf. Communications (ICC'94)*, May 1996, pp. 645–649.

[16] G. Kaplan, S. Shamai (Shitz), and Y. Kofman, "On the design and selection of convolutional codes for an uninterleaved, bursty Rician channel," *IEEE Trans. Commun.*, vol. 43, pp. 2914–2921, Dec. 1995.

[17] J. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.

[18] M. D. Trott, "The algebraic structure of trellis codes," Ph.D. dissertation, MIT, Aug. 1992.

[19] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 56–67, Jan. 1982.

[20] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, "Impact of diversity reception on fading channels with coded modulation. Part I: Coherent detection," to be published in *IEEE Trans. Commun.*, vol. 45, pp. 563–572, May 1997.

[21] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.

[22] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*, vol. 8, May 1990.

[23] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[24] A. J. Viterbi, J. K. Wolf, E. Zehavi, and R. Padovani, "A pragmatic approach to trellis-coded modulation," *IEEE Commun. Mag.*, vol. 27, pp. 11–19, July 1989.

[25] L. F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, vol. 41, pp. 1439–1449, Oct. 1993.

[26] E. Zehavi, "8-PSK trellis codes for a rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.

Authorized licensed use limited to: Fish & Richardson PC. Downloaded on May 01,2022 at 15:14:37 UTC from IEEE Xplore. Restrictions apply.