TOWARDS THE CAPACITY OF NONCOHERENT ORTHOGONAL MODULATION: BICM-ID FOR TURBO CODED NFSK

Matthew C. Valenti
West Virginia University
Morgantown, WV
mvalenti@csee.wvu.edu

Ewald Hueffmeier and Bob Bogusch
Mission Research Corporation
Monterey, CA
ewald@mrcmry.com
rlbogusch@earthlink.net

John Fryer
Applied Data Trends
Huntsville, AL
jfryer@hiwaay.net

ABSTRACT

This paper* investigates coded communications using noncoherent orthogonal modulation and capacity-approaching binary channel codes. The focus is on the interface between the decoder and demodulator, which is critical for large modulation order M. While standard receivers for bit-interleaved coded-modulation (BICM) segregate the operations of demodulation and decoding, we follow the recently proposed paradigm of BICM with iterative decoding (BICM-ID) to approximate joint demodulation and decoding through the iterative exchange of soft information between demodulator and decoder. We focus our attention on the derivation of the soft-input/soft-output SISO demodulator for noncoherent M-FSK, and derive a log-domain SISO demodulator suitable for channels with uniform phase and fading amplitudes that are either known or constant (i.e. AWGN). Simulation results are shown for $M = 2, 4, 16,$ and 64 using the well-known UMTS turbo code. It is found that feeding soft information from the decoder back to the demodulator improves performance by between 0.7-0.9 dB for 16-ary and 64-ary NFSK in AWGN and Rayleigh fading.

INTRODUCTION

Receivers used in wireless military communication systems must often operate in the presence of phase uncertainty. If the channel coherence time is sufficiently long, the lack of phase information can be alleviated by using pilot symbol assisted modulation (PSAM) or differential phase shift keying (DPSK) [1]. However, if the channel is subject to extreme Doppler spread, the channel coherence time might be on the order of a single symbol. For such systems, orthogonal modulation with noncoherent detection, as typified by noncoherent frequency shift keying (NFSK), is a natural choice. Even if the channel coherence time is relatively long, an inexpensive system using low-cost oscillators might not be able to maintain phase coherence over a sufficiently large

number of symbols. Again, noncoherent orthogonal modulation is a viable candidate for such systems. For instance, IEEE 802.15 task group 4 has proposed the use of orthogonal 16-ary modulation with noncoherent detection for low rate wireless personal area networks in the 2.4 GHz ISM band [2].

Turbo codes are capable of approaching within 0.5 dB of the channel capacity of binary NFSK in AWGN [3] and within 0.7 dB of capacity in Rayleigh flat-fading [4]. However, the $\mathcal{E}_b/N_o$ required to achieve capacity using binary NFSK is quite large (in excess of 6.7 dB). One of the benefits of using orthogonal modulation is that it allows for a tradeoff between energy-efficiency and bandwidth [5]. By using a higher order modulation, the required $\mathcal{E}_b/N_o$ is decreased. In systems that are limited by energy rather than bandwidth (e.g. many military systems and sensor network applications), larger values of M (the number of orthogonal signals in the signal set) are desired.

A pragmatic approach to coding for M-ary modulation with $M > 2$ is *bit interleaved coded modulation* (BICM) [6]. With BICM, a *binary* channel code is created, interleaved bit-wise, and then passed to a M-ary modulator. While slightly inferior to trellis-coded modulation (TCM) in AWGN, BICM is actually superior to TCM in fading because it maximizes the Hamming distance, which is more important than squared-Euclidian distance in fading [6]. With conventional BICM receivers, a demodulator produces soft estimates for each code bit which is then decoded with a standard soft-input decoder. Performance of BICM can be improved by feeding soft information from the decoder back to the demodulator, a process known as *bit interleaved coded modulation with iterative decoding* (BICM-ID) [7]. BICM-ID has been considered for several types of M-ary modulation, including 8-PSK [8, 9] and QAM [10]. Most work on BICM-ID to date has focused only on two-dimensional modulation formats; BICM-ID used with M-ary orthogonal modulation has been virtually ignored. One notable exception is found in [11], which proposes an iterative noncoherent demodulator and turbo decoder for turbo coded orthogonal modulation. The receiver proposed in [11] uses the BICM-ID concept, although it never explicitly uses this term. Unfortunately, the one-page limit of the conference precluded a detailed expo-
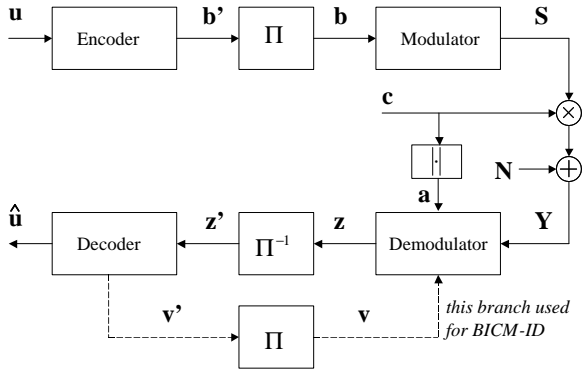
Figure 1: System model.

sition. Furthermore, [11] only indicates a performance improvement of 0.1 dB in the waterfall region of the turbo code when using 64-FSK. Results shown here (table I) indicate a gain of 0.87 dB for AWGN and Rayleigh fading for the same modulation order, indicating that perhaps the system proposed in [11] was suboptimal.

In this paper, we consider iterative demodulation and decoding for turbo-coded noncoherent orthogonal modulation. We derive the optimal soft-input/soft-output (SISO) symbol-by-symbol demodulator for noncoherent orthogonal modulation. A log-domain expression is given that permits efficient and numerically stable implementation of the SISO demodulator for NFSK. Simulation results are given for the rate $R = 1/3$ UMTS turbo code, and these results are compared to the corresponding channel capacity [12].

Before proceeding further, let us stipulate some notational conventions. Bold lowercase letters will be used to denote vectors, e.g. $\mathbf{x}$, and bold uppercase will be used for matrices, e.g. $\mathbf{X}$. All vectors are row-vectors, but can be transposed into column vectors, e.g. $\mathbf{x}^T$. Vector elements are plain lowercase letters with subscripts beginning at zero, e.g. $\mathbf{x} = [x_0, x_1, ..., x_{M-1}]$. Matrices are represented as a row of column vectors, e.g. $\mathbf{X} = [\mathbf{x}_0^T, \mathbf{x}_1^T, ..., \mathbf{x}_{N-1}^T]$. The function $p(\cdot)$ represents the probability of an event, a probability density function, or a probability mass function with the context clearly dependent upon the argument.

SYSTEM MODEL

The discrete-time system model is shown in Fig. 1. A vector $\mathbf{u} \in \{0, 1\}^K$ of message bits is passed through a binary encoder to produce a codeword $\mathbf{b}' \in \{0, 1\}^N$ which is interleaved by a permutation matrix $\mathbf{\Pi}$ to produce the bit-interleaved codeword $\mathbf{b} = \mathbf{b}'\mathbf{\Pi}$. The bit-interleaved codeword is then passed through a M-ary orthogonal modulator to produce the $M \times L$ matrix of symbols $\mathbf{S} = [\mathbf{s}_0^T, ..., \mathbf{s}_{L-1}^T]$ where $L = \lceil N/\log_2 M \rceil$. Each column of $\mathbf{S}$ represents one M-ary symbol and is represented as an elementary vector $\mathbf{e}_m$

comprised of all zeros except for a one in the mth position.

Assume that arbitrary symbol $\mathbf{s}$ is transmitted. Without loss of generality, assume that the first $\mu = \log_2 M$ bits in $\mathbf{b}$ are gathered to form the symbol, i.e. $\mathbf{s} \Leftrightarrow \{b_0, ..., b_{\mu-1}\}$. With orthogonal modulation, the mapping of code bits to symbols is unimportant since the symbols are equidistant, and thus a natural mapping suffices. In this case, the symbol $\mathbf{s} = \mathbf{e}_m \in \{\mathbf{e}_0, ..., \mathbf{e}_{M-1}\}$ where the index

$$m = \sum_{k=0}^{\mu-1} b_k 2^k. \tag{1}$$

The coded symbol stream passes through a frequency-nonselective channel with complex fading amplitudes $\mathbf{c} \in \mathbb{C}^L$. The ith fading coefficient can be represented as $c_i = a_i \exp\{\theta_i \sqrt{-1}\}$, where a_i and θ_i are the real-valued amplitude and phase, respectively. In general, the c_i's could have any distribution, but in the following discussion we focus on two cases: (1) AWGN: $a_i = 1$ and the θ_i's are i.i.d. uniform on $[0, 2\pi)$; and (2) Rayleigh Fading: the c_i's are i.i.d. zero-mean complex Gaussian with a variance of $1/2$ in both the real and imaginary directions and thus the a_i's are Rayleigh and θ_i's are uniform. The ith fading coefficient c_i is multiplied by the ith symbol $\mathbf{s}_i$ and the result is added to the ith column $\mathbf{n}_i^T$ of the noise matrix $\mathbf{N} \in \mathbb{C}^{M \times L}$ which contains uncorrelated zero-mean complex Gaussian noise samples with variance $\sigma^2 = 1/(2\mathcal{E}_s/N_o)$ in both directions (N_o is the one-sided noise spectral density). The energy per coded symbol $\mathcal{E}_s$ is related to the energy per message bit $\mathcal{E}_b$ by $\mathcal{E}_s = K\mathcal{E}_b/L$. The received complex symbols $\mathbf{Y} = [(c_0\mathbf{s}_0^T + \mathbf{n}_0^T), ..., (c_{L-1}\mathbf{s}_{L-1}^T + \mathbf{n}_{L-1}^T)]$ are then passed to the receiver.

The input to the conventional noncoherent BICM demodulator is the matrix of received symbols $\mathbf{Y}$, an estimate of the average symbol signal-to-noise ratio $\mathcal{E}_s/N_o$, and estimates of the fading amplitudes $\mathbf{a} = [a_0, ..., a_{L-1}]$. Because the receiver knows $\mathbf{a}$, it is said to have *channel state information* (CSI). An estimate of $\mathcal{E}_s/N_o$ can be computed using the method in [4], while estimates for $\mathbf{a}$ can be found in correlated fading by using a Wiener filter matched to the channel's autocorrelation, possibly aided by periodically inserted pilot symbols [13].

In addition, the BICM-ID demodulator has available extrinsic information $\mathbf{v}$ produced by the soft-output decoder, which is used by the demodulator as a priori estimates of the likelihood of the code bits. The demodulator interprets elements of $\mathbf{v}$ as the log-likelihood ratio

$$v_k = \log \frac{\hat{p}_k}{1 - \hat{p}_k}, \tag{2}$$

where $\hat{p}_k$ is the decoder's estimate of the probability that $b_k = 1$; because this is extrinsic information, it is produced using information about all code bits *other than* b_k.

The SISO demodulator works on a symbol-by-symbol basis, producing soft information for all μ code bits associated with a particular symbol, using only those portions of the demodulator inputs $\mathbf{Y}, \mathbf{v}$, and $\mathbf{a}$ that pertain to that symbol. For an arbitrary symbol $\mathbf{s}$, let $\mathbf{y}$ be the received signal vector, c the complex channel gain, $a = |c|$ the channel amplitude, and $\tilde{\mathbf{v}} = [v_0, ..., v_{\mu-1}]$ be the portion of $\mathbf{v}$ that corresponds to this symbol. The SISO demodulator computes a log-likelihood ratio for each code bit b_k, in the form

$$\lambda_k = \log \frac{p(b_k = 1|\mathbf{y}, \tilde{\mathbf{v}}, a)}{p(b_k = 0|\mathbf{y}, \tilde{\mathbf{v}}, a)}. \tag{3}$$

If a noniterative BICM receiver [6] is used, the conditioning on $\tilde{\mathbf{v}}$ is removed (represented in Fig. 1 by removing the $\mathbf{v}$ input at the bottom of the demodulator block). As will be shown in (6), the LLR can be decomposed into $\lambda_k = z_k + v_k$, where z_k is *extrinsic information*. To prevent the harmful positive feedback of probabilities, only z_k is passed to the channel decoder.

The extrinsic information at the output of the demodulator is deinterleaved and the resulting sequence $\mathbf{z}' = \mathbf{z}\mathbf{\Pi}^{-1}$ is passed to a soft-input decoder which produces a vector $\hat{\mathbf{u}}$ containing hard estimates of the message bits. The BICM-ID receiver requires that the decoder produces extrinsic information $\mathbf{v}'$ of the code bits, which is reinterleaved to form the input $\mathbf{v}$ to the demodulator. The details of the soft-output decoder will not be discussed here, as it has already been treated extensively in the literature. To obtain our simulation results, we used the soft-input/soft-output (SISO) algorithm of [14] implemented in the log-domain [15, 16].

The BICM-ID receiver iterates between demodulation and decoding, with the reliability of the exchanged extrinsic information improved after each half-iteration. Note that we have placed no requirements on the type of code. If the code itself does not require iterative decoding, for instance if it is a conventional convolutional code, then it is natural for there to be a single iteration of decoding for every iteration of demodulation. For non-iteratively decoded codes, the iterative nature of the BICM-ID receiver substantially increases the complexity of the system, which grows linearly in the number of iterations. For instance, [7] uses a convolutional code and three iterations of BICM-ID decoding, thereby tripling the complexity relative to the non-iterative BICM receiver. On the other hand, if the code must be iteratively decoded, BICM-ID does not impose a heavy burden on the overall system complexity. This is because the receiver already must iterate for the sake of decoding, and thus updating the soft demodulator statistics between each decoder iteration is less of a burden than requiring entire iterations solely for the sake of BICM-ID. Complexity can be reduced further by allowing the decoder to execute several local iterations of decoding before updating the soft demodulator metric (this is done in [11] which executes 5 local iterations of turbo decoding for each of 2 global iterations of BICM-ID).

SOFT NFSK DEMODULATOR

Now let us turn our attention to the calculation of (3). The first step is to partition the two probabilities in (3) over the set of symbols,

$$\lambda_k = \log \frac{\sum_{i \in S_k^{(1)}} p(\mathbf{s}_i|\mathbf{y}, \tilde{\mathbf{v}}, a)}{\sum_{i \in S_k^{(0)}} p(\mathbf{s}_i|\mathbf{y}, \tilde{\mathbf{v}}, a)}, \tag{4}$$

where the set $S_k^{(1)}$ contains the indices of all symbols labelled with $b_k = 1$, and $S_k^{(0)}$ contains the indices of all symbols labelled with $b_k = 0$. To compute the summands in (4), first apply Bayes' rule

$$p(\mathbf{s}_i|\mathbf{y}, \tilde{\mathbf{v}}, a) = \frac{p(\mathbf{y}|\mathbf{s}_i, \tilde{\mathbf{v}}, a)p(\mathbf{s}_i, \tilde{\mathbf{v}}, a)}{p(\mathbf{y}, \tilde{\mathbf{v}}, a)} \tag{5}$$

The extrinsic information $\tilde{\mathbf{v}}$ pertaining to this symbol is generated by the decoder using information pertaining to symbols other than this one. As a consequence, $\tilde{\mathbf{v}}$ is independent of a, $\mathbf{s}_i$, and $\mathbf{y}$ and thus $p(\mathbf{y}|\mathbf{s}_i, \tilde{\mathbf{v}}, a) = p(\mathbf{y}|\mathbf{s}_i, a)$. Since $\{b_k = b\}$ and $S_k^{(b)}$ are *equivalent events*, $p(\mathbf{s}_i, \tilde{\mathbf{v}}, a) = p(\mathbf{s}_i, \tilde{\mathbf{v}}, a, b_k = b)$ for $i \in S_k^{(b)}$. Furthermore, a is independent of $\mathbf{s}_i, \tilde{\mathbf{v}}$, and b_k, thus $p(\mathbf{s}_i, \tilde{\mathbf{v}}, a, b_k = b) = p(a)p(\mathbf{s}_i, \tilde{\mathbf{v}}, b_k = b)$. From the definition of conditional probability $p(\mathbf{s}_i, \tilde{\mathbf{v}}, b_k = b) = p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = b)p(\tilde{\mathbf{v}}, b_k = b) = p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = b)p(b_k = b|\tilde{\mathbf{v}})p(\tilde{\mathbf{v}})$. Gathering all these terms, we get for $i \in S_k^{(b)}$

$$p(\mathbf{s}_i|\mathbf{y}, \tilde{\mathbf{v}}, a) = \frac{p(\mathbf{y}|\mathbf{s}_i, a)p(a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = b)p(b_k = b|\tilde{\mathbf{v}})p(\tilde{\mathbf{v}})}{p(\mathbf{y}, \tilde{\mathbf{v}}, a)}$$

Inserting this back into (4), cancelling common terms, and taking the $p(b_k = b|\tilde{\mathbf{v}})$ term out of the summation yields

$$\begin{aligned} \lambda_k &= \log \frac{p(b_k = 1|\tilde{\mathbf{v}}) \sum_{i \in S_k^{(1)}} p(\mathbf{y}|\mathbf{s}_i, a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = 1)}{p(b_k = 0|\tilde{\mathbf{v}}) \sum_{i \in S_k^{(0)}} p(\mathbf{y}|\mathbf{s}_i, a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = 0)} \\ &= v_k + \log \frac{\sum_{i \in S_k^{(1)}} p(\mathbf{y}|\mathbf{s}_i, a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = 1)}{\sum_{i \in S_k^{(0)}} p(\mathbf{y}|\mathbf{s}_i, a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = 0)}, \end{aligned} \tag{6}$$

where the last equality follows from (2). The demodulator outputs the extrinsic information $z_k = \lambda_k - v_k$, or

$$z_k = \log \frac{\sum_{i \in S_k^{(1)}} p(\mathbf{y}|\mathbf{s}_i, a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = 1)}{\sum_{i \in S_k^{(0)}} p(\mathbf{y}|\mathbf{s}_i, a)p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k = 0)} \tag{7}$$

This expression clearly delineates the contribution of the channel observation, which influences only the $p(\mathbf{y}|\mathbf{s}_i, a)$ term, and the contribution of the a priori information passed to the demodulator from the decoder, which affects only the $p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k)$ term.

When the demodulator does not use a priori information from the decoder, as in a conventional BICM receiver, then

$p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k) = p(\mathbf{s}_i|b_k) = 2/M$, and since these terms are all equal, they cancel out in (7). But now consider what happens if the demodulator has available an estimate $\hat{p}_k$ of the probability that $b_k = 1$, which can be found from the a priori LLR input $\tilde{\mathbf{v}}$. Let symbol $\mathbf{s}_i$ be labelled by $\{b_0^{(i)}, ..., b_{\mu-1}^{(i)}\}$. Under the assumption of independent code bits (achieved by proper interleaving), the symbol probability is

$$
\begin{aligned}
p(\mathbf{s}_i|\tilde{\mathbf{v}}, b_k) &= \prod_{\substack{j=0 \\ j\neq k}}^{\mu-1} p(b_j = b_j^{(i)}|v_j) \\
&= \prod_{\substack{j=0 \\ j\neq k}}^{\mu-1} b_j^{(i)} \hat{p}_j + (1-\hat{p}_j)(1-b_j^{(i)}) \\
&= \prod_{\substack{j=0 \\ j\neq k}}^{\mu-1} \frac{b_j^{(i)} e^{v_j} + (1-b_j^{(i)})}{1 + e^{v_j}} \\
&= \prod_{\substack{j=0 \\ j\neq k}}^{\mu-1} \frac{e^{b_j^{(i)} v_j}}{1 + e^{v_j}}
\end{aligned} \tag{8}
$$

where the third line follows from the second from (2).

Now consider the $p(\mathbf{y}|\mathbf{s}_i, a)$ term in (7). Since $\mathbf{y} = c\mathbf{s}+\mathbf{n}$, $\mathbf{y}$ is complex Gaussian with mean $c\mathbf{e}_i$ when conditioned on symbol $\mathbf{s}_i$ and complex fading coefficient c. Thus its conditional pdf is, from [5],

$$
p(\mathbf{y}|\mathbf{s}_i, c) =
$$

$$
\left(\frac{\mathcal{E}_s}{\pi N_o}\right)^M \exp\left\{-\frac{\mathcal{E}_s}{N_o}\left[|y_i - c|^2 + \sum_{\substack{j=0 \\ j\neq i}}^{M-1} |y_j|^2\right]\right\}. \tag{9}
$$

Because the demodulator is noncoherent, the phase θ of c is unknown and thus (9) must be marginalized with respect to θ. Assuming that θ is uniform, as also shown in [5],

$$
p(\mathbf{y}|\mathbf{s}_i, a) = \int_0^{2\pi} p(\theta)(\mathbf{y}|\mathbf{s}_i, c)d\theta = \tag{10}
$$

$$
\left(\frac{\mathcal{E}_s}{\pi N_o}\right)^M \exp\left\{-\frac{\mathcal{E}_s}{N_o}\left[a^2 + \sum_{j=0}^{M-1} |y_j|^2\right]\right\} I_0\left(\frac{2\mathcal{E}_s a|y_i|}{N_o}\right),
$$

where $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind. If the distribution of θ is not uniform, as in Rician fading, then (10) needs to be calculated with respect to the corresponding $p(\theta)$ [12].

The soft demodulator output z_k is found by substituting (8) and (10) into (7). However, this operation can be simplified by exploiting the fact that (7) contains a ratio of probabilities and thus many terms may cancel. For instance, the $(1 + e^{v_j})$ in the denominator of (8) will cancel in the ratio and can be dropped. Furthermore, all terms except for the Bessel function will cancel in (10). Thus,

$$
z_k = \log \frac{\displaystyle\sum_{i \in S_k^{(1)}} I_0\left(\frac{2\mathcal{E}_s a|y_i|}{N_o}\right) \prod_{\substack{j=0 \\ j\neq k}}^{\mu-1} \exp\left(b_j^{(i)} v_j\right)}{\displaystyle\sum_{i \in S_k^{(0)}} I_0\left(\frac{2\mathcal{E}_s a|y_i|}{N_o}\right) \prod_{\substack{j=0 \\ j\neq k}}^{\mu-1} \exp\left(b_j^{(i)} v_j\right)}. \tag{11}
$$

This expression can be further simplified by using the max-star operator as defined in [16],

$$
\max_i {}^* \{x_i\} = \log\left\{\sum_i e^{x_i}\right\}, \tag{12}
$$

where the pairwise max-star operator is defined as $\max*(x,y) = \max(x,y) + \log(1 + e^{-|x-y|}) = \max(x,y) + f_c(|x-y|)$ and multiple arguments imply a recursion of pairwise operations, i.e. $\max*(x,y,z) = \max*(x, \max*(y,z))$. In terms of $\max*$, (11) becomes

$$
\begin{aligned}
z_k = &\max_{i \in S_k^{(1)}} {}^* \left[\log I_0\left(\frac{2\mathcal{E}_s a|y_i|}{N_o}\right) + \sum_{\substack{j=0 \\ j\neq k}}^{\mu-1} b_j^{(i)} v_j\right] \\
&-\max_{i \in S_k^{(0)}} {}^* \left[\log I_0\left(\frac{2\mathcal{E}_s a|y_i|}{N_o}\right) + \sum_{\substack{j=0 \\ j\neq k}}^{\mu-1} b_j^{(i)} v_j\right].
\end{aligned} \tag{13}
$$

Much of the computational complexity of the above expression lies in the calculation of several nonlinear functions. Note, however, that the logarithm and Bessel function always appear together as $\log[I_o(\cdot)]$ and so this combined function can be implemented as a single table look-up. A piecewise linear approximation for this nonlinear function is given in [3]. Also note that the arguments of the $\log[I_o(\cdot)]$ operator are all channel observations that do not change after any iteration of decoding (only $\mathbf{v}$ changes after each demodulator iteration). Thus, the $\log[I_o(\cdot)]$ calculation need only be performed prior to the first iteration. Another frequently computed nonlinear function is the correction function $f_c(z) = \log(1 + e^{-z})$ that must be calculated by each pairwise max-star operation. This function can be implemented using the linear approximation in [17]. Alternatively, by noting that $\max*(x,y) \approx \max(x,y)$, each $\max*$ operator in (13) could be replaced with max, although this imposes a performance penalty (we observed a loss of approximately 0.5 dB over a wide range of parameters). The additional complexity per-bit-iteration of using BICM-ID is

at worst $M(\mu - 1)$ additions and $M - 2$ pairwise max* (or max) operations. This complexity could be further reduced by exploiting the structure in the natural mapping given by (1). Note that the $\log[I_o(\cdot)]$ must be computed even for BICM, and since it is only computed once in BICM-ID, it imposes no additional burden. Given the complexity of the turbo decoder and the nonlinear $\log[I_o(\cdot)]$ operation, the additional complexity of BICM-ID is quite manageable, especially for reasonable values of M.

SIMULATION RESULTS

To illustrate the effectiveness of the proposed BICM-ID technique for M-ary NFSK, we conducted an extensive set of simulations. For the channel code, the full-length turbo code from the UMTS specification was used [18]. This code has length $(N, K) = (15354, 5114)$, including tail bits, and is (roughly) rate $R = 1/3$. The BICM interleaver Π was implemented as a μ by L block interleaver, with bits written into the interleaver row-wise and read out column-wise. Several other interleaver designs were also examined, including s-random interleavers and interleavers designed according to the three rules in [9]. We found, however, that performance was not significantly influenced by interleaver design, presumably due to the fact that the turbo code already contains its own internal interleaver.

We considered both AWGN and fully-interleaved Rayleigh flat-fading. In all cases, it is assumed that the average value of $\mathcal{E}_b/N_o$ is known at the receiver and in the fading case, the fading amplitudes **a** are known. Four values of the modulation order M were considered, $M = 2, 4, 16$, and 64. For $M > 2$, both BICM and BICM-ID were considered (for $M = 2$, BICM-ID is equivalent to BICM). In each case, 16 iterations of BICM-ID decoding were performed (with a single local iteration of turbo decoding for each global iteration of BICM-ID). For every data point, the simulation ran until at least 30 frame errors were recorded.

Fig. 2 shows BER performance for M=16 in AWGN as a function of the number of iterations. This plot shows curves for both BICM (solid lines) and BICM-ID (dashed lines). From right to left, the performance after iterations 1,2,3,4,5,10, and 16 are shown (the performance after 1 iteration is identical for both systems). The curves indicate that the performance of BICM-ID after 3 iterations is always better than the performance of BICM after all 16 iterations. This implies that, although BICM-ID is marginally more complex per iteration than BICM, a system using BICM-ID can actually be much less complex than BICM because it can achieve the same performance by running far fewer iterations.

Fig. 3 shows performance in AWGN after all 16 iterations for all four values of M considered. For each $M > 2$ a pair of curves are shown, one for BICM and the other for BICM-ID.
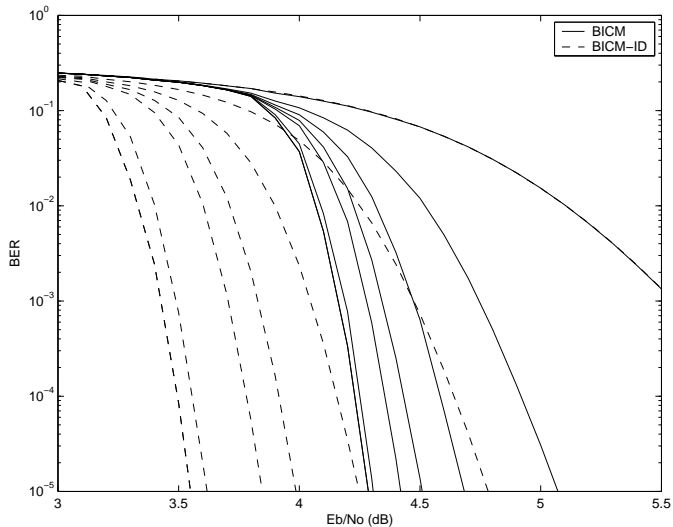

Figure 2: BER vs. $\mathcal{E}_b/No$ for the UMTS turbo code and 16-ary NFSK using both BICM (solid line) and BICM-ID (dashed line) in AWGN. From right to left, the curves show performance after 1, 2, 3, 4, 5, 10, and 16 iterations.
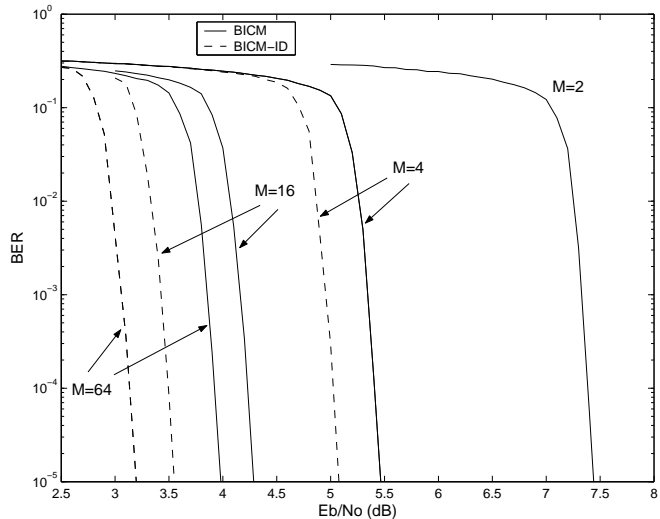

Figure 3: BER vs. $\mathcal{E}_b/No$ for the UMTS turbo code using M-ary NFSK and both BICM (solid line) and BICM-ID (dashed line) in AWGN after 16 iterations for different modulation orders.
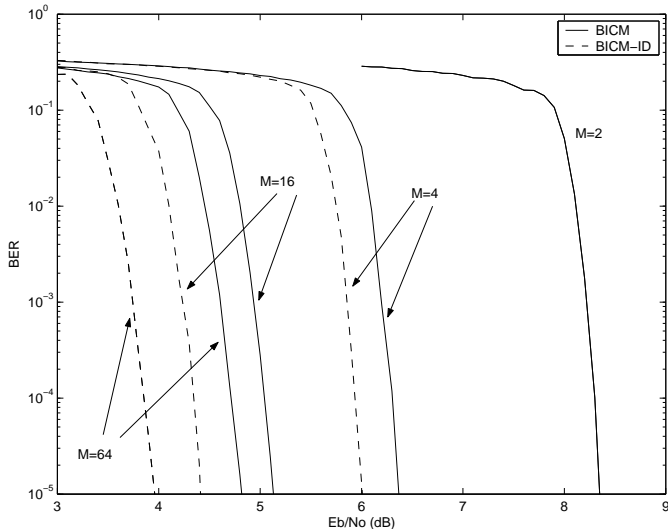
Figure 4: BER vs. $\mathcal{E}_b/No$ for the UMTS turbo code using M-ary NFSK and both BICM (solid line) and BICM-ID (dashed line) in fully-interleaved Rayleigh flat fading after 16 iterations for different modulation orders.

Type	M	BICM	BICM-ID	Capacity
AWGN	2	7.44 dB	N/A	6.86 dB
	4	5.47 dB	5.08 dB	4.35 dB
	16	4.29 dB	3.55 dB	2.30 dB
	64	3.98 dB	3.20 dB	1.37 dB
Rayleigh Fading	2	8.35 dB	N/A	7.55 dB
	4	6.37 dB	6.00 dB	5.01 dB
	16	5.13 dB	4.41 dB	2.91 dB
	64	4.82 dB	3.96 dB	1.94 dB

Table 1: Minimum $\mathcal{E}_b/N_o$ required to achieve a BER of 10^{-5} using the full-length UMTS turbo code, M-ary noncoherent FSK, and either BICM or the proposed BICM-ID technique. The corresponding Shannon capacity is also given.

The corresponding curves for the fully-interleaved Rayleigh flat fading channel is shown in Fig. 4.

In Table I, we list the value of $\mathcal{E}_b/N_o$ required to achieve a BER of 10^{-5} using the UMTS turbo code for each modulation order in both AWGN and Rayleigh fading. The table indicates the required value of $\mathcal{E}_b/N_o$ when using BICM and BICM-ID. Also, the Shannon capacity for M-ary NFSK is given. These values for capacity were found by solving the multidimensional integration given in [12] by applying the Monte Carlo integration technique of [19].

The dB gain due to using BICM-ID increased with M, with gains between 0.37 and 0.39 dB for $M = 4$, between 0.72 and 0.74 dB for $M = 16$, and between 0.78 and 0.86 for $M = 64$. The gains in fading and AWGN were comparable. Despite the impressive gains due to using BICM-ID, performance is still further away from capacity than it is with binary NFSK. In AWGN, the gap to capacity increases from 0.58 dB for $M = 2$ to 0.73, 1.25, and 1.83 dB for $M = 4, 16$, and 64, respectively. In Rayleigh fading, the gap increases from 0.8 dB for $M = 2$ to 1, 1.5, and 2 dB for $M = 4, 16$, and 64, respectively. As the performance improvement due to BICM-ID increases with M, so does the gap to capacity, suggesting that further improvements to this process are possible.

CONCLUSIONS

The combination of binary turbo codes and binary noncoherent orthogonal modulation can approach the Shannon capacity to within a fraction of a dB. When the same binary turbo code is bit-interleaved, modulated with M-ary orthogonal modulation, then demodulated noncoherently and finally decoded, the gap to Shannon capacity increases with M. This gap can be partially closed by approximating joint demodulation and decoding with a simpler algorithm that iterates between a SISO demodulator and decoder. For instance, the performance of 64-ary NFSK in AWGN is improved by 0.78 dB relative to segregated demodulation and decoding. In this case performance is still 1.83 dB from capacity, suggesting that further improvements should be possible. One source of improvement is through careful interleaver design, but our attempts at improved interleaver design proved fruitless when we used the UMTS turbo code. Any attempt to improve the interleaving should jointly consider the turbo code's internal interleaver with the global BICM interleaver.

In our work, we assumed that the average $\mathcal{E}_b/N_o$ is known at the receiver, as our prior experience and reports in the literature indicate that estimating $\mathcal{E}_b/N_o$ over a large block of data is not difficult and turbo decoding is robust against slight SNR estimation errors [20]. Nevertheless, the problem of SNR estimation should be considered more thoroughly before the proposed technique can be fielded. Likewise, our results for Rayleigh fading assumed that the amplitude estimates can be perfectly estimated. While this is difficult when the channel coherence time is short, accurate estimates can be achieved if the channel coherence time is sufficiently long by periodically transmitting pilot symbols or tones [21]. Alternatively, the receiver could operate without channel state information by also marginalizing out a in (10). A more sophisticated blind receiver could actually switch between operating with and without CSI. During the first few iterations, the receiver does not know the fading amplitudes and must operate without CSI. However, as the decoder begins to resolve the data, the fading amplitudes could be estimated. Then the estimated fading amplitudes could provide CSI for future iterations. A hybrid receiver could operate without CSI for some symbols and with CSI for only those symbols whose amplitudes have been reliably estimated.

REFERENCES

[1] R. R. Chen, R. Koetter, U. Madhow, and D. Agrawal, "Joint noncoherent demodulation and decoing for the block fading channel: A practical framework for approaching Shannon capacity," *IEEE Trans. Commun.*, vol. 51, pp. 1676–1689, Oct. 2003.

[2] LAN/MAN Standards Committee of the IEEE Computer Society, "Draft standard for part 15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low rate wireless personal area networks (LR-WPANs)," *Draft P802.15.4/D18*, pp. 27–52, Feb. 2003.

[3] E. K. Hall and S. G. Wilson, "Turbo codes for noncoherent channels," in *Proc., IEEE GLOBECOM, Communication Theory Mini-Conference*, (Phoenix, AZ), pp. 66–70, Nov. 1997.

[4] A. Ramesh, A. Chockalingam, and L. B. Milstein, "Performance of noncoherent turbo detection on Rayleigh fading channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, (San Antonio, TX), pp. 1193–1198, Nov. 2001.

[5] J. Proakis, *Digital Communications.* New York, NY: McGraw-Hill, Inc., fourth ed., 2001.

[6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[7] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 1, pp. 169–171, Nov. 1997.

[8] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit-interleaving and iterative decoding," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 715–724, Apr. 1999.

[9] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8-PSK signaling," *IEEE Trans. Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.

[10] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation of BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 944–957, May 2001.

[11] P. C. P. Liang and W. E. Stark, "Algorithm for joint decoding of turbo codes and M-ary orthogonal modulation," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, (Sorrento, Italy), p. 191, June 2000.

[12] W. E. Stark, "Capacity and cutoff rate of noncoherent FSK with nonselective Rician fading," *IEEE Trans. Commun.*, vol. 33, pp. 1153–1159, Nov. 1985.

[13] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1697–1705, Sept. 2001.

[14] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, pp. 22–24, Jan. 1997.

[15] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," *European Trans. on Telecommun.*, vol. 8, pp. 119–125, Mar./Apr. 1997.

[16] A. J. Viterbi, "An intuitive justification and a simplified implemetation of the MAP decoder for convolutional codes," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 260–264, Feb. 1998.

[17] M. C. Valenti and J. Sun, "The UMTS turbo code and an efficient decoder implementation suitable for software defined radios," *Int. J. Wireless Info. Networks*, vol. 8, pp. 203–216, Oct. 2001.

[18] European Telecommunications Standards Institute, "Universal mobile telecommunications system (UMTS): Multiplexing and channel coding (FDD)," *3GPP TS 125.212 version 3.4.0*, pp. 14–20, Sept. 23 2000.

[19] S. J. MacMullan and O. M. Collins, "The capacity of orthogonal and bi-orthogonal codes on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1217–1232, May 1998.

[20] T. A. Summers and S. G. Wilson, "SNR mismatch and online estimation in turbo decoding," *IEEE Trans. Commun.*, vol. 46, pp. 421–423, April 1998.

[21] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, vol. 40, pp. 686–693, Nov. 1991.