Turbo-NFSK: Iterative Estimation, Noncoherent Demodulation, and Decoding for Fast Fading Channels

Shi Cheng and Matthew C. Valenti West Virginia University Morgantown, WV {shic,mvalenti}@csee.wvu.edu

ABSTRACT

This paper^{*} considers the problem of communicating over fast fading channels, where the channel coherence time is only on the order of a few symbols. Since the fading is too fast for coherent reception, we employ M-ary frequency shift keying with soft noncoherent demodulation. Information is encoded by a binary turbo code. To improve performance, the soft demodulator and decoder work cooperatively through the iterative exchange of extrinsic information. During each iteration, the receiver estimates the channel state information (CSI), here defined to be the average received signal energy and noise spectral density for each block of symbols. The channel estimator uses the Expectation Maximization (EM) algorithm and exploits extrinsic information fed from the decoder. Simulation results show that for 16-NFSK in block independent Ravleigh fading, performance can be within 0.6 dB of the perfect CSI case by using blocks as small as 4 symbols.

INTRODUCTION

Bit interleaved coded modulation (BICM) [1] combines binary forward error correcting coding with M-ary modulation. It has become a standard method for signaling over fading channels, forming the basis of third generation cellular and 802.11a/g wireless networks. The performance of BICM can be improved by feeding back a priori information (in the form of bit likelihoods) from the decoder back to the demodulator. Such iterative demodulation and decoding schemes were independently developed by ten Brink [2], Benedetto et al [3], and Li and Ritcey [4]. The latter reference terms this technique bit interleaved coded modulation with iterative decoding (BICM-ID).

When signaling over a fading channel, one of two possible techniques is typically used. The first option is to periodically insert pilot symbols into the transmitted signal, and then leverage these pilot symbols to perform coherent detection [5]. This is effective only if the fading is sufficiently slow and the transmit and receive oscillators relatively stable. AnDon Torrieri U.S. Army Research Laboratory Adelphi, MD dtorrieri@arl.army.mil

other option is to use orthogonal signaling and noncoherent reception. This is more appropriate when either dealing with fast fading or when the oscillators are not stable enough, for instance in frequency hopping applications. The focus of this paper is on the second option.

A benefit of orthogonal signaling, such as frequency shift keying $(FSK)^{\dagger}$ is that it allows bandwidth efficiency to be traded for energy efficiency. If a binary code is combined with nonbinary FSK, then BICM-ID can improve performance, as shown in [6].

In [6], the performance of turbo coded FSK using BICM-ID was shown under the assumption that the channel fading amplitude was known perfectly at the receiver. However, in practice, this amplitude is not known a priori and therefore must be estimated. This paper extends the work of [6] by including the process of channel estimation into the receiver structure. To facilitate the development of a pragmatic estimator, it is assumed that the channel experiences block fading, that is, blocks of N consecutive FSK symbols are attenuated by the same channel gain (though they could possibly experience different phase shifts). Aside from this block fading condition, the estimator makes no assumptions regarding the statistics of the channel and, in fact, estimates each block independently from the other blocks. The estimator itself is derived using the expectation maximization (EM) algorithm [7], which iteratively finds the maximum likelihood (ML) estimate, even though an explicit form is not readily achievable when extrinsic information is fed back to the estimator from the decoder.

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. $\mathbf{X} = [\mathbf{x}_0^T, \mathbf{x}_1^T, ..., \mathbf{x}_{M-1}^T]$. The function

^{*}This work was sponsored by the Xenotran Corporation, Glen Burnie, MD.

 $^{^{\}dagger}\mathrm{In}$ this paper, orthogonal modulation and FSK are used interchangeably.



Figure 1: System model.

 $p(\cdot)$ represents the probability of an event, a probability density function, or a probability mass function with the context clearly dependent on 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 $M \times N_f$ matrix of symbols $\mathbf{S} = [\mathbf{s}_0^T, ..., \mathbf{s}_{N_f-1}^T]$ where $N_f = \lceil n/\log_2 M \rceil$. 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 m^{th} position.

Let the set of $\mu = \log_2 M$ code bits that label symbol \mathbf{s}_i be represented as $\{b_0^{(i)}, ..., b_{\mu-1}^{(i)}\}$. With orthogonal modulation and flat fading that is constant during each symbol period, the manner in which the code bits are mapped to symbols is unimportant since the symbols are equidistant, and thus a natural mapping suffices. In this case, $\mathbf{s}_i = \mathbf{e}_m \in \{\mathbf{e}_0, ..., \mathbf{e}_{M-1}\}$ where

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

The modulated symbol \mathbf{s}_i , which has unit energy, is then scaled by a factor of $\sqrt{\mathcal{E}_s}$, so that the transmitted symbol is $\mathbf{x}_i = \sqrt{\mathcal{E}_s} \mathbf{s}_i$, and the transmitted matrix of scaled symbols is $\mathbf{X} = [\mathbf{x}_0^T, ..., \mathbf{x}_{N_f-1}^T]$.

The modulated symbol stream passes through a frequencynonselective block fading channel, with the fading coefficient constant within each block, and independent over different blocks. In the following discussion, we assume $N_f = NL$, where N is the number of symbols per block, and L is the number of blocks per codeword. The ℓ^{th} block fading coefficient can be represented as $c_{\ell} = a_{\ell} \exp\{\theta_{\ell}\sqrt{-1}\}$, where a_{ℓ} and θ_{ℓ} are the real-valued amplitude and phase, respectively. Thus, the received signal of the ℓ^{th} block can be represented as

$$\mathbf{Y}_{\ell} = c_{\ell} \mathbf{X}_{\ell} + \mathbf{N}_{\ell}, \qquad (2)$$

where \mathbf{X}_{ℓ} and \mathbf{N}_{ℓ} consist of columns ℓN through $(\ell+1)N-1$ of \mathbf{X} and \mathbf{N} , respectively, and \mathbf{N} is a $M \times NL$ noise matrix whose elements are independently and identically distributed (i.i.d.) complex Gaussian variables whose real and imaginary components each have zero mean and variance $N_0/2$.

In (2), the $(k,i)^{th}$ entry of \mathbf{Y}_{ℓ} , is denoted y_{ki} , where $k = 0, 1, 2, \dots, M-1$, $i = 0, 1, 2, \dots, N-1$. When the i^{th} symbol is received, it is passed through a filter bank with M matched filters, one for each FSK tone. Since the receiver is noncoherent, each matched filter produces a complex quantity. The element y_{ki} is then the output of the k^{th} matched filter during the i^{th} symbol period.

The receiver processes the matrix of received symbols Y and produces estimates $\hat{\mathbf{u}}$ of the data bits. The receiver is decomposed into a demodulator and a decoder, each of which are implemented using the soft-input soft-output algorithm of [3] and separated by appropriate interleaving/deinterleaving. The demodulator processes the matrix of received symbols \mathbf{Y} to produce extrinsic information \mathbf{z} , which is deinterleaved and passed to the decoder. While a coherent receiver will have knowledge of the complex fading gains \mathbf{c} , the noncoherent receiver will at best only know the fading amplitudes $\mathbf{a} = [a_0, ..., a_{L-1}]$. When **a** is available, the system is said to have perfect *channel state information* (CSI), while when it is not available the system has no channel state information (NCSI). Details of the demodulator can be found in [6][8]. Briefly, the output of the demodulator when a, \mathcal{E}_s , and N_o are known is

$$z_{k} = \max_{i \in S_{k}^{(1)}} \left[\log I_{0} \left(\frac{2a_{\ell} \sqrt{\mathcal{E}_{s}} |y_{i}|}{N_{o}} \right) + \sum_{\substack{j=0\\j \neq k}}^{\mu-1} b_{j}^{(i)} v_{j} \right] \\ -\max_{i \in S_{k}^{(0)}} \left[\log I_{0} \left(\frac{2a_{\ell} \sqrt{\mathcal{E}_{s}} |y_{i}|}{N_{o}} \right) + \sum_{\substack{j=0\\j \neq k}}^{\mu-1} b_{j}^{(i)} v_{j} \right],$$
(3)

where $S_k^{(1)}$ is the set of indices of those symbols whose k^{th} bit is labeled with a 1, $S_k^{(0)}$ is the set of indices of those symbols whose k^{th} bit is labeled with a 0, and the pairwise max-star operator is defined as $\max(x, y) = \max(x, y) + \log(1 + e^{-|x-y|})$. Multiple arguments imply a recursion of pairwise operations, i.e. $\max(x, y, z) = \max(x, \max(y, z))$. In the above, **v** is the extrinsic information fed from the decoder back to the demodulator, in log-likelihood ratio (LLR) form.

CHANNEL ESTIMATOR WITH KNOWN TRANSMITTED SEQUENCE

When \mathcal{E}_s , N_0 , and **a** are not known to the receiver, (3) can not be used directly. Instead, these parameters are first estimated and then the estimates are used in place of the actual values in (3). For each $M \times N$ received block \mathbf{Y}_{ℓ} , the estimator will generate estimates of the received amplitude $a_{\ell}\sqrt{\mathcal{E}_s}$ and the noise spectral density N_0 . Note that since we are not assuming the presence of a perfect automatic gain control (AGC) unit, we estimate these quantities separately rather than following the usual convention of estimating the ratio \mathcal{E}_s/N_o .

For robustness purposes, the estimator makes no assumptions regarding the distribution of the quantities to be estimated, nor does it make any assumptions regarding correlation from block to block. While the estimator performance could be improved by exploiting inter-block correlation and a priori knowledge of the channel statistics (for instance, that the **a**'s are Rician distributed), this leads to a far less robust estimator and increases complexity since parameters such as fading rate and Rician K-factor must also be estimated. If the assumptions about the channel prove false, or if the channel parameters change too quickly to be tracked, then the performance of such an estimator will be far worse than the one proposed here.

Each block is processed in an identical fashion. The estimator directly uses the channel observation for the ℓ^{th} block, \mathbf{Y}_{ℓ} , while the observations of the other blocks are used indirectly through feedback of extrinsic information from the decoder. Since the form of the estimation algorithm is the same for each block, in the following discussion we let \mathbf{Y} represent a generic $M \times N$ block and a the corresponding fading amplitude. While we could have indicated the block index ℓ , such notation is not necessary and only obscures the results.

Following [9], we can represent the probability density function of the noncoherent received signals as

$$p(y_{ki}|\mathbf{s}_{i} = \mathbf{e}_{j})$$

$$\frac{1}{\pi N_{0}} \exp\left(-\frac{|y_{ki}|^{2} + a^{2} \mathcal{E}_{s} \delta_{kj}}{N_{0}}\right) I_{0}\left(\frac{2a\sqrt{\mathcal{E}_{s}}|y_{ki}|\delta_{kj}}{N_{0}}\right),$$
(4)

where δ_{kj} is the Kroneker delta function ($\delta_{kj} = 1$ if k = j, otherwise $\delta_{kj} = 0$).

From (4), we can easily form the pdf of the i^{th} symbol as

$$p(\mathbf{y}_i|\mathbf{s}_i = \mathbf{e}_j) = (5)$$

$$\left(\frac{1}{\pi N_0}\right)^M \exp\left(-\frac{\sum_{k=0}^{M-1} |y_{ki}|^2 + a^2 \mathcal{E}_s}{N_0}\right) I_0\left(\frac{2a\sqrt{\mathcal{E}_s} |y_{ji}|}{N_0}\right)$$

The log-likelihood function with respect to \mathcal{E}_s and N_0 can be

represented as

$$L = \ln [f(\mathbf{Y}|A, B, \mathbf{d})]$$

= $-MN \ln A - \frac{C}{A} - \frac{NB^2}{4A} + \sum_{i=0}^{N-1} \ln I_0 \left(\frac{B |y_{d_i i}|}{A}\right),$
(6)

where $C = \sum_{k,i} |y_{ki}|^2$, $A = N_0$ and $B = 2a\sqrt{\mathcal{E}_s}$. Here, **d** is an $N \times 1$ vector, and each element d_i indicates that the i^{th} actual transmitted symbol \mathbf{s}_i is \mathbf{e}_{d_i} .

In order to find the maximum-likelihood (ML) estimate, we need to take the derivatives with respect to A and B, and find the solution to $\frac{\partial L}{\partial A} = 0$ and $\frac{\partial L}{\partial B} = 0$. After some manipulation to simplify the estimates, the following ML estimates are obtained:

$$\hat{A} = \frac{1}{MN} \left(C - \frac{N\hat{B}^2}{4} \right) \tag{7}$$

$$\hat{B} = \frac{2}{N} \sum_{i=0}^{N-1} |y_{d_i i}| F\left(\frac{4MN\hat{B}|y_{d_i i}|}{4C - N\hat{B}^2}\right), \quad (8)$$

where $F(x) = I_1(x)/I_0(x)$ with I_{ν} being the modified Bessel function of the first kind and order ν . Equations (7) and (8) produce the ML estimate of A and B when the exact transmitted sequence is known to the receiver. However, the sequence is unknown in a practical receiver whose objective is to detect and decode the transmitted sequence.

EM ESTIMATOR

The representation of (6) includes the unknown parameter \mathbf{d} . In order to find the ML solution, we need to sum over all possibilities of \mathbf{d} , by using the extrinsic information feedback from the decoder. This results in

$$L = \ln \prod_{i=0}^{N-1} f(\mathbf{y}_i | A, B)$$

=
$$\sum_{i=0}^{N-1} \left[\ln \sum_{k=0}^{M-1} f(\mathbf{y}_i | A, B, d_i = k) p(d_i = k) \right].$$
 (9)

Even though the expression can be broken down to the product of independent symbols, the argument of the log function still contains the summation of M terms, and therefore a direct solution is too complex to be practical.

• The EM algorithm is an appropriate tool for finding the solution to this problem with reasonable complexity. In order to apply the EM algorithm, we first define the complete data set as $\{\mathbf{Y}, \mathbf{d}\}$. Then, the log-likelihood function is formed for the complete data set, resulting in

$$\tilde{L} = \ln [f(\mathbf{Y}, \mathbf{d}|A, B)]
= \ln [f(\mathbf{Y}|A, B, \mathbf{d})] + \ln p(\mathbf{d})
\sim -MN \ln A - \frac{C}{A} - \frac{NB^2}{4A} + \sum_{i=0}^{N-1} \ln I_0 \left(\frac{B |y_{d_i i}|}{A}\right).$$
(10)

The log-likelihood functions of (10) and (6) for the complete and incomplete data, respectively, are identical because we assume that the *a priori* distribution of **d** is independent of *A* and *B* and, hence, may be omitted. The EM algorithm is then performed in the following two steps:

$$E - step:$$

$$Q(A, B | \hat{A}^{(l)}, \hat{B}^{(l)}) = E \left[\tilde{L}(A, B) | \mathbf{Y}, \hat{A}^{(l)}, \hat{B}^{(l)} \right] (11)$$

$$\mathbf{M} - step:$$

$$\hat{A}^{(l+1)}, \hat{B}^{(l+1)} = \arg \max Q(A, B | \hat{A}^{(l)}, \hat{B}^{(l)}) (12)$$

To form the expectation of the E-step, we take the expectation over **d** conditioned on $\hat{A}^{(l)}, \hat{B}^{(l)}$ and **Y**,

$$Q(A, B | \hat{A}^{(l)}, \hat{B}^{(l)})$$

$$= -MN \ln A - \frac{C}{A} - \frac{NB^2}{4A}$$

$$+ \sum_{i=0}^{N-1} E_{d_i | \mathbf{y}_i, \hat{A}^{(l)}, \hat{B}^{(l)}} \left[\ln I_0 \left(\frac{B | y_{d_i i} |}{A} \right) \right]$$

$$= -MN \ln A - \frac{C}{A} - \frac{NB^2}{4A}$$

$$+ \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \tilde{p}_{ki} \ln I_0 \left(\frac{B | y_{ki} |}{A} \right), \quad (13)$$

where

$$\tilde{p}_{ki} = p(d_i = k | \mathbf{y}_i, \hat{A}^{(l)}, \hat{B}^{(l)})
= \frac{p(\mathbf{y}_i | d_i = k, \hat{A}^{(l)}, \hat{B}^{(l)}) p(d_i = k)}{p(\mathbf{y}_i | \hat{A}^{(l)}, \hat{B}^{(l)})}$$
(14)

Applying (5), we obtain

$$\tilde{p}_{ki} = \alpha_i^{(l)} I_0 \left(\frac{\hat{B}^{(l)} |y_{ki}|}{\hat{A}^{(l)}} \right) p(d_i = k).$$
(15)

where $\alpha_i^{(l)}$ is the normalization factor forcing $\sum_{k=0}^{M-1} \tilde{p}_{ki} = 1$; thus,

$$\alpha_i^{(l)} = \frac{1}{\sum_{k=0}^{M-1} I_0\left(\frac{\hat{B}^{(l)}|y_{ki}|}{\hat{A}^{(l)}}\right) p(d_i = k)}.$$
 (16)

The probability $p(d_i = k)$ can be formed by combining the bit probabilities fed back from the decoder (after first converting from LLR domain to probability domain) [8]:

$$p(d_i = k) = \prod_{j=0}^{\mu-1} \frac{\exp\left[b_j^{(k)} v_{i,j}\right]}{1 + \exp\left[v_{i,j}\right]},$$
(17)

where the *a priori* log-likelihood ratios corresponding to the μ bits associated with the i^{th} symbol are $v_{i,0}, v_{i,1}, \cdots, v_{i,\mu-1}$. Estimates of *A* and *B* are then found using

$$\hat{A}^{(l+1)} = \frac{1}{MN} \left(C - \frac{N(\hat{B}^{(l+1)})^2}{4} \right)$$
(18)
$$\hat{B}^{(l+1)} = \frac{2}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \tilde{p}_{ki} |y_{ki}| F\left(\frac{4MN\hat{B}^{(l+1)}|y_{ki}|}{4C - N(\hat{B}^{(l+1)})^2}\right).$$

(19)

Since a closed form solution to (19) is difficult to obtain, an iterative method needs to be implemented within each EM iteration. This is satisfied by the recursion

$$B_{n+1}^{(l+1)} = \frac{2}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \tilde{p}_{ki} |y_{ki}| F\left(\frac{4MNB_n^{(l+1)}|y_{ki}|}{4C - N(B_n^{(l+1)})^2}\right),$$
(20)

where the superscript denotes the EM iteration, while the subscript denotes the iteration for the recursion (20). To initialize, the initial value of $B_0^{(l+1)}$ can be simply set to $\hat{B}^{(l)}$ from the previous EM iteration. The initial values used in (15) are set to

$$\hat{B}^{(0)} = \frac{2}{N} \sum_{i=0}^{N-1} \max_{k} |y_{ki}|$$
(21)

$$\hat{A}^{(0)} = \frac{1}{MN} \left(C - \frac{N(\hat{B}^{(0)})^2}{4} \right).$$
(22)

which are only based on the channel observations \mathbf{y} , regardless of the channel SNR.

Thus, this two-fold iterative process works as follows: First, the initial values $\hat{A}^{(0)}$ and $\hat{B}^{(0)}$ are computed. Based on the initial values, the probabilities $\{\tilde{p}_{ki}\}$ are calculated according to (15), (16), and (17). The iterative method (20) can then be implemented to find $\hat{B}^{(1)}$, with an initial setting $B_0^{(1)} = \hat{B}^{(0)}$. Using $\hat{B}^{(1)}$ and (18), $\hat{A}^{(1)}$ can be easily found in a single step, which finalizes the first EM iteration. The second EM iteration then starts by recalculating \tilde{p}_{ki} based on the newly acquired $\hat{A}^{(1)}$ and $\hat{B}^{(1)}$, and the remaining steps are similar to the first EM iteration. The whole algorithm terminates when $\hat{A}^{(l)}$ or $\hat{B}^{(l)}$ converges to some fixed value.

Now we will look into the convergence behavior of (20). Let the function $\tilde{G}(B_n^{(l+1)})$ denote the right-hand side of (20). During our simulations, it was found that the $\tilde{G}(B)$ exhibits concavity. Since F(t) approaches 1 asymptotically when t goes to infinity,

$$\lim_{B \to \sqrt{\frac{4C}{N}}} \tilde{G}(B) = \frac{2}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \tilde{p}_{ki} |y_{ki}| \\ \leq \frac{2}{N} \sum_{i=0}^{N-1} \max_{k} |y_{ki}|$$
(23)

Applying the Cauchy-Schwarz inequality for inner products and then the definition of C, we obtain

$$\lim_{B \to \sqrt{\frac{4C}{N}}} \tilde{G}(B) < \sqrt{\frac{4C}{N}}.$$
(24)

These inequalities and the concavity of $\tilde{G}(B)$ indicate that the point $(B_0^{(1)}, \tilde{G}(B_0^{(1)}))$ is to the upper-right of $(\hat{B}^{(1)}, \hat{B}^{(1)})$, so that the iteration of (20) guarantees $B^{(1)}$ gradually approaches $\hat{B}^{(1)}$. The simulation shows that the ML estimates converge quickly, in less than 10 EM iterations. In practice, for a BICM-ID system, (22) and (21) are only used to select the initial values for the first pass. From the second pass, the estimators can make use of the ML estimates calculated during the previous pass.

SIMULATION RESULTS

To illustrate the performance of the proposed estimator, a set of simulations were run. The simulated system uses the turbo code from the cdma2000 specification [10]. The turbo code was a rate-1/2 code with 1530 input bits. As the cdma2000 standard requires 12 coded tail bits, the length of each code word is 2(1530) + 12 = 3072 bits. In the case that 16-FSK modulation is used, each FSK symbol conveys 4 code bits, and the number of FSK symbols per code word is 768. The receiver executed up to 20 iterations (an early halting routine stopped the iterations once the data was correctly decoded).

Fig. 2 and 3 show the simulation results under independent Rayleigh fading. Blocks of N consecutive symbols undergo the same fading amplitude, while the fading amplitudes are i.i.d. Rayleigh from block to block. The phase of each symbol is uniformly distributed, though the distribution of the phases does not affect the performance of the proposed system.

In Fig. 2, only 16-FSK modulation was simulated, while several values of block length N were considered: N =1, 4, 8, 16, 32, and 64. For each value of N two curves are shown, one for the case that the CSI is perfectly known by the receiver, and the other for the case that the CSI is estimated by the proposed EM-based estimator. The perfect CSI case serves as a benchmark to compare against the performance of the estimator and therefore always exhibits better performance. As can be seen from the figure, performance of the perfect CSI case always improves with decreasing N. This is because as N decreases there is a corresponding increase in the number L of independent fading blocks per code word, which translates into an increase in diversity. On the other hand, the performance of the estimated CSI case does not always improve with decreasing N. In the simulated case, the estimated CSI case improves with decreasing N up until N = 4 but then gets worse for the N = 1 case. This is because at the same time that diversity is increasing, the length of each block is reduced resulting in a poorer estimate of the



Figure 2: Performance of the rate 1/2 cdma2000 turbo code with K = 1530 input bits and 16-FSK modulation for several block lengths (N). Blocks are i.i.d. Rayleigh, and the performance with perfect CSI (dashed lines) is compared against the performance using the proposed estimator (solid lines).

fading coefficient. This behavior can also be seen by the tendency for the gap between the perfect CSI and estimated CSI cases to widen with decreasing N. This is a common issue for fading channels; faster fading improves performance by providing more diversity while simultaneously degrading performance by providing the estimator with fewer samples. For N = 64 the proposed estimator is less than 0.1 dB worse than the perfect CSI case, while for N = 4 the estimator is 0.6 dB worse, which is still quite acceptable. For N = 1performance is rather poor, being nearly 2 dB away from the perfect CSI case. When N = 1 it would be better to use an alternative form of the demodulator that does not require CSI, see for instance [11] and [6].

Fig. 3 considers the influence of different FSK alphabet sizes when the data rate and block duration are held constant at 24 bits per block. In other words, a single block contains 24 binary FSK symbols, 12 4-FSK symbols, 6 16-FSK symbols or 4 64-FSK symbols. The channel code is the same as the one in Fig. 2, and there are 128 blocks per codeword. As the modulation alphabet size M goes up, the required \mathcal{E}_b/N_o decreases for both the case with perfect CSI and the case with the EM-based estimator. The simulation result shows that the loss due to the proposed estimator is about 0.5 dB, which varies little with the alphabet size M as long as the block duration remains constant at 24 bits per block.

Fig. 4 shows simulation results for i.i.d. Rician block fading with a 10 dB Rician K factor. While this figure shows the same tendency for the gap between perfect and estimated CSI to widen with decreasing block length, the gap is smaller



Figure 3: Performance of the rate 1/2 cdma2000 turbo code with K = 1530 input bits and blocks that are 24 bits long for several alphabet sizes (*M*). Blocks are i.i.d. Rayleigh, and the performance with CSI (dashed lines) is compared against the performance using the proposed estimator (solid lines).

compared with the Rayleigh fading case. For example, the gap decreases from 0.6 dB to 0.4 dB when there is 4 16-FSK symbols per block. This can be attributed to Rician fading being less severe than Rayleigh.

In the previous curves, the relationship between performance and block length was complicated by the tradeoff between diversity and estimation error. As the block length decreased, the diversity improved, but the estimation error got worse. To better illuminate the effect of block length on estimator performance, Fig. 5 shows simulation results for the same cdma2000 turbo code and 16-FSK in an unfaded, AWGN channel. Because the channel is unfaded, the fading amplitude is unity for the entire codeword. For the perfect CSI case, there is just a single curve. The other curves on this figure show the performance of the EM-based estimator for different estimator block lengths. While the \mathcal{E}_s and N_o are constant for the whole codeword, the estimator runs under the assumption that they are fixed for only Nconsecutive symbols. When N = 1, the performance of the estimator is about 0.7 dB away from the perfect CSI case, and the gap becomes smaller as N increases. This reason is that more accurate estimation can be achieved through larger block size, and unlike the fading case, increasing Ndoes not decrease diversity. When N = 4, the gap is only about 0.3 dB. When N increases to 32, the proposed estimator has almost the same performance as the perfect CSI case.



Figure 4: Performance of the rate 1/2 cdma2000 turbo code with K = 1530 input bits and 16-FSK modulation as a function of block length N. Blocks are i.i.d. Rician with K = 10dB, and the performance with perfect CSI (dashed lines) is compared against the performance using the proposed estimator (solid lines).



Figure 5: Performance of the rate 1/2 cdma2000 turbo code with K = 1530 input bits and 16-FSK modulation in AWGN. The performance with the proposed estimator (solid lines) using different block sizes (N) is compared against reception with known \mathcal{E}_s and N_o (dashed line).

CONCLUSIONS

The turbo-NFSK system is a robust noncoherent system that has been shown to withstand the severe channel conditions of fast fading, unknown fading attenuation, unknown fading statistics, and unknown noise-power spectral density. The channel-state estimator is based on the Expectation Maximization algorithm and exploits extrinsic information produced after each decoding iteration of the turbo code. Each updated channel-state estimate is applied to the next decoder iteration. Simulation results indicate that if the fading coherence time exceeds four channel symbols, then the performance is close to what could be obtained with perfect channel-state information.

REFERENCES

- G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.
- [2] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, pp. 806–808, May 13, 1999.
- [3] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks," *European Trans. on Telecommun.*, vol. 9, pp. 155–172, Mar.-Apr. 1998.
- [4] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 1, pp. 169–171, Nov. 1997.

- [5] 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.
- [6] M. C. Valenti and S. Cheng, "Iterative demodulation and decoding of turbo coded *M*-ary noncoherent orthogonal modulation," *IEEE J. Select. Areas Commun.*, Sept. 2005.
- [7] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extentions*. Willey, 1997.
- [8] M. C. Valenti, E. Hueffmeier, B. Bogusch, and J. Fryer, "Towards the capacity of noncoherent orthogonal modulation: BICM-ID for turbo coded NFSK," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, (Monterey, CA), Nov. 2004.
- [9] D. Torrieri, Principles of Spread-Spectrum Communication Systems. Springer, 2004.
- Third Generation Partnership Project 2 (3GPP2), "Physical layer standard for cdma2000 spread spectrum systems, release C," 3GPP2 C.S0002-C Version 1.0, pp. 115–122, May 28 2002.
- [11] W. E. Stark, "Capacity and cutoff rate of noncoherent FSK with nonselective Rician fading," *IEEE Trans. Commun.*, vol. 33, pp. 1153–1159, Nov. 1985.