An Iterative Multiuser Decoder for Near-Capacity Communications

Michael Moher, *Member, IEEE*

***Abstract*—The combination of forward error correction (FEC) coding and random interleaving is shown to overcome the limitations of multiuser detectors/decoders when the user cross correlations are high. In particular, one can asymptotically achieve single-user performance in a highly correlated multiuser system. In addition, an optimal iterative multiuser detector is derived from iterative techniques for cross-entropy minimization. A practical suboptimal implementation of this algorithm is presented, and simulations demonstrate that, even with highly correlated users, it achieves optimal asymptotic efficiency. The effects of the theoretical limits on channel capacity are evident in many of the simulation results. The complexity of the suboptimal algorithm is approximately $(O(2^K)+O(2^\kappa))$ per bit per iteration where K is the number of users and κ is the code constraint length.**

***Index Terms*—Iterative decoding, minimum cross entropy, multiuser decoding, multiuser detection.**

I. Introduction

MULTIUSER detection has received considerable attention recently with its potential to improve system capacity and alleviate some technical requirements of code-division multiple-access (CDMA) systems, such as power control. Many algorithms for performing multiuser detection have been put forth. These range from the high-complexity optimum detectors for asynchronous systems [1] to many forms of suboptimum lower complexity detectors including linear [2]–[4] and nonlinear [5], [6]. In many scenarios these detectors work well and provide close to optimum performance.

The interest in CMDA has rekindled interest in multiuser decoders and recent results in this area include [7]–[13] and [26]. In [9] it is demonstrated that the optimum detector for an asynchronous CDMA system employing forward error correction (FEC) coding combines the trellises of both the asynchronous detector and the FEC code. The result is a time-varying trellis with a complexity per bit that is approximately $O(2^{\kappa K})$, where K is the number of users and κ is the code constraint length. This complexity makes the use of the optimal decoding detector prohibitive for even small systems.

A limitation of these multiuser detectors/decoders is that they degrade as the correlation between users increases. This limits the applicability of these techniques to situations where the correlation between users is intentionally kept low, such as CDMA. In this paper we investigate how to extend these techniques to the case where the correlations are high, such as frequency-division multiple-access (FDMA) with a high degree of channel overlap. An optimal iterative multiuser decoder is derived based on cross-entropy minimization techniques [24], [25] and a practical implementation of this algorithm is presented. The resulting suboptimal algorithm is similar to one that has been independently discovered by Reed *et al.* [18] for CDMA.

In Section II we review notation used for the multiuser detection problem and illustrate some of the limits of conventional multiuser detectors/decoders. It is shown how these limits can be overcome. In Section III it is shown how techniques similar to those applied to the detection of parallel concatenated codes (turbo codes) [15], [16] can also be applied to detecting parallel users, that is, multiuser decoding (MUD). In Section IV simulation results are provided that confirm the theoretical results of the previous two sections. In Section V a further analysis of the simulation results is performed.

Paper approved by S. S. Pietrobon, the Editor for Coding Theory and Techniques of the IEEE Communications Society. Manuscript received July 28, 1997; revised October 29, 1997 and February 10, 1998. This paper was presented in part at the IEEE International Symposium on Information Theory (ISIT'97), Ulm, Germany, June 1997.

The author is with the Communications Research Centre, Ottawa, Ont. K2H 8S2, Canada (e-mail: michael.moher@crc.doc.ca).

Publisher Item Identifier S 0090-6778(98)05171-X.

II. Multiuser Detection

In the following sections we assume K bit-synchronous users, each transmitting a block of N bits. For bit period i of user k, the corresponding transmitted bit will be represented by $b_k(i)$. The complete sequence of the kth user's bits will be represented by the vector $\mathbf{b}_k = (b_k(1), \cdots, b_k(N)) \in \{-1,+1\}^N$ and the vector of the K-user bits at time i is represented by $\mathbf{b}(i) = (b_1(i), \cdots, b_K(i))^T$. The matrix of all users' bits is represented by $\mathbf{b} \in \{-1,+1\}^{K\times N}$, where $\mathbf{b}_k$ is a row of this matrix. The same notation is extended to other discrete-time quantities.

A. Multiuser Model

The model for multiple access communications, illustrated in Fig. 1, assumes that there are K users simultaneously accessing the same communications channel using modulating waveforms $s_k(t) \in \Re(-\infty,\infty)$, $k=1,\cdots,K$, with symbol period T. The noiseless input to the receiver is

$$S(\mathbf{b},t) = \sum_{i=1}^{N}\sum_{k=1}^{K} b_k(i)\sqrt{w_k}s_k(t-iT-\tau_k). \quad (1)$$

The parameters $\{\tau_k\}$ represent the relative transmission delays and, for the synchronous case, $\tau_1 = \cdots = \tau_k = 0$. The modulating waveforms $s_k(t)$ are assumed to be normalized

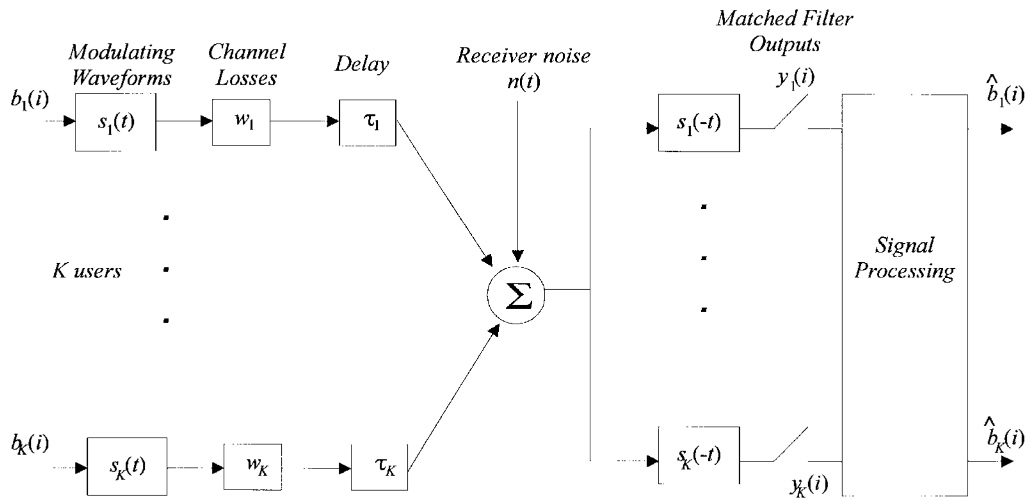

Fig. 1. Illustration of multiuser communication model.

to unit energy. The relative received amplitude levels of the different users are characterized by the positive parameters $\{\sqrt{w_k}\}$. For the case of equal power users, $w_1 = w_2 = \cdots = w_K = 1$. With an additive noise channel, the corresponding received signal is

$$r(t) = S(\mathbf{b}, t) + n(t). \tag{2}$$

It is assumed that the noise process $n(t)$ is a zero-mean complex white Gaussian process with spectral density N_0. This model can be easily extended to include complex signaling and can be applied to a variety of multiple-access systems including FDMA and CDMA. With FDMA, the modulating waveforms are the carrier frequencies with appropriate pulse shaping.[1] With CDMA, the modulating waveforms correspond to the spreading codes assigned to each user.

It is assumed that the modulating waveform of each user is known at the receiver, and that the K-user coherent receiver locks to the signaling interval and phase of each active user. In a highly correlated multiuser environment it is recognized that constructing a coherent receiver is a challenging problem, but this is not dealt with in this paper. If the additive noise is a white Gaussian random process, it can be shown [1] that a set of sufficient statistics for maximum-likelihood detection of the data is the set of matched-filter outputs

$$y_k(i) = \int_{-\infty}^{\infty} \mathrm{Re}[r(t)s_k(t - iT - \tau_k)]\, dt \tag{3}$$

where $y_k(i)$ is the output of a filter matched to the kth modulating waveform. Define the klth entry in the cross-correlation matrix of the user waveforms $H(i) \in \Re^{K\times K}$ by

$$(H(i))_{kl} = \int_{-\infty}^{\infty} s_k(t - iT - \tau_k)s_l(t - \tau_l)\, dt. \tag{4}$$

For the synchronous case, it is assumed that the cross correlation between modulating waveforms in adjacent symbol intervals is zero, that is, $H(i) = 0$ for $|i| > 0$. The latter is usually forced by assuming that the modulating waveform is zero outside the symbol interval, i.e., $s_k(t) \in \Re[0, T]$. Under these assumptions, letting $H = H(0)$, an equivalent discrete-time representation of this system is

$$\mathbf{y}(i) = HW^{1/2}\mathbf{b}(i) + \mathbf{n}(i) \quad \mathbf{y}(i), \mathbf{b}(i), \mathbf{n}(i) \in \Re^K \tag{5}$$

where the vector $\mathbf{n}(i)$ is a set of zero-mean correlated noise samples with

$$n_k(i) = \int_{-\infty}^{\infty} \mathrm{Re}[n(t)s_k(t - iT - \tau_k)]\, dt \tag{6}$$

and W is the diagonal matrix with nonzero elements $\{w_k\}$. Since it is assumed that the noise $n(t)$ is white, it follows from (4) and (6) that $E[\mathbf{n}(i)\mathbf{n}(j)^T] = \sigma^2 H\delta(i - j)$, where $\sigma^2 = N_0/2T$.

As a reference, the optimum approach performs an exhaustive search to determine that $\mathbf{b}(i)$

$$\hat{\mathbf{b}}(i) = \arg\min_{\mathbf{x}\in\{-1,+1\}^K} \{(\mathbf{y}(i) - HW^{1/2}\mathbf{x})^T H^{-1}(\mathbf{y}(i) - HW^{1/2}\mathbf{x})\}. \tag{7}$$

The conventional detector takes the sign of the bits at the output of the matched filter, that is, $\hat{\mathbf{b}}(i) = \mathrm{sign}\{\mathbf{y}(i)\}$, where the $\mathrm{sign}(\cdot)$ operator is applied on an element-by-element basis. The conventional approach is much less complex than the optimum approach but clearly relies on low cross correlations, that is, H being approximately diagonal, for good performance.

B. Detector Efficiency

The *efficiency* of a detector is a measure of the power efficiency of a multiuser detector relative to a single-user detector operating at the same bit-error rate (BER). It allows a simple comparison between detectors of different types. The *asymptotic efficiency* is defined as [2]

$$\eta_k = \sup\left\{0 \le r \le 1\colon \lim_{\sigma\to 0} \frac{P_K^k(\sigma)}{P_1^k(\sigma/\sqrt{r})} < \infty\right\} \tag{8}$$

where $P_K^k(\sigma)$ is the error rate of the kth user in the multiuser system and $P_1^k(\sigma)$ is the error rate of the kth user in a single-user system with the same noise variance σ^2. For the optimum

[1] With FDMA, the modulating waveforms are complex, in general, with a complex phase that may vary from one symbol interval to the next.

detector on a Gaussian channel, the asymptotic error rate is determined by the minimum Euclidean distance between any pair of K-dimensional sequences $\mathbf{b}^1$ and $\mathbf{b}^2$ that differ for the kth user. This distance can be expressed as

$$d_k^2 = \min_{\substack{\mathbf{b}^1,\mathbf{b}^2 \\ \mathbf{b}_k^1 \neq \mathbf{b}_k^2}} \|S(\mathbf{b}^1,t) - S(\mathbf{b}^2,t)\|^2 \tag{9}$$

where $\|(\cdot)\|^2 = \int_{-\infty}^{\infty} |(\cdot)|^2 \, dt$. For the synchronous Gaussian channel, performance at low noise is dominated by the minimum distance error, and the asymptotic efficiency of the optimum detector reduces to the ratio of the scaled minimum distances. Evaluating (9) for the multiuser and single-user minimum distances, in a manner similar to [3], and taking the ratio[2] gives

$$\eta_k = \frac{1}{w_k d_{\text{free}}} \min_{\varepsilon_k \neq 0} \sum_{n=1}^{N} \varepsilon(n)^T W^{1/2} H W^{1/2} \varepsilon(n) \tag{10}$$

where the minimum is over all K-dimensional error sequences $\varepsilon \in \{-1, 0, +1\}^{K \times N}$, and the error sequence ε_k corresponding to the kth user is nonzero. The quantity d_{free} is the Hamming distance of the minimum distance single-user error event ($d_{\text{free}} = 1$ for uncoded) and $w_k d_{\text{free}}$ is the corresponding Euclidean distance. Asymptotic efficiency is analogous to asymptotic coding gain, as it represents the gain (loss) relative to an uncoded (single-user) system as the signal-to-noise ratio (SNR) becomes large.

C. Example Uncoded Efficiencies

To illustrate these concepts, we define the K-symmetric channel that is characterized by a cross-correlation matrix

$$H = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \rho & \rho & 1 & \cdots \\ \cdots & \cdots & \rho & 1 \end{bmatrix} \tag{11}$$

and a channel gain matrix of $W^{1/2} = I$. That is, all users have the same cross correlation between their modulating waveforms and suffer the same propagation losses. Design of the modulating waveforms of the K users to produce (11) is, in general, not possible without some bandwidth expansion. There are some common examples. The case $\rho = 0$ corresponds to orthogonal signaling, for example, using the orthogonal Hadamard codes in a CDMA system or orthogonal frequencies in an FDMA system. Using different cyclic shifts of the same M sequence, as the modulating waveforms of the users, corresponds to the case $\rho = -1/(2^n - 1)$. Using the same modulating waveform for each user corresponds to $\rho = 1$, and the bandwidth expansion factor is one. The K-symmetric channel provides insight into the behavior of multiuser detectors as a function of the number of users because it allows the analytical calculation of the asymptotic efficiency as a function of the cross-correlation parameter ρ and the number of users K [14]. In Fig. 2 the asymptotic efficiencies of the conventional, the decorrelating [2], [3], and the optimal detectors are compared for 2, 5, and 10 users and positive cross correlations ρ for the K-symmetric channel.

[2] A factor of four has been removed from both the numerator and denominator.

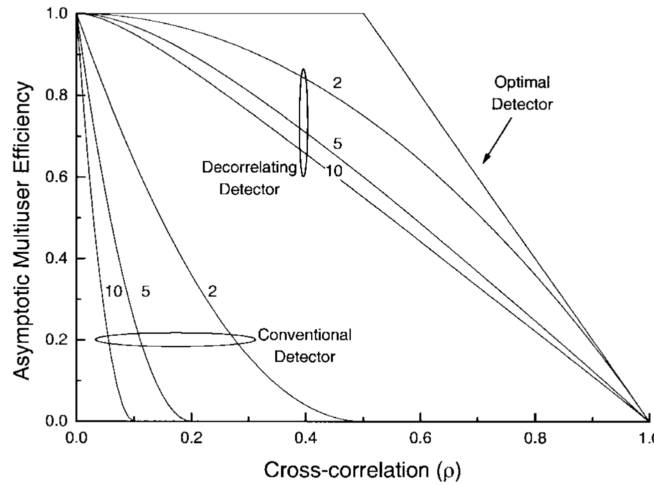

Fig. 2. Asymptotic efficiencies of conventional, decorrelating, and optimum detectors as a function of the number of users (K) and the cross-correlation parameter ρ.

For this channel, the asymptotic efficiency of the optimal detector is independent of the number of users for positive ρ. Both the conventional and decorrelating detectors degrade with the number of users. The asymptotic efficiency of the conventional detector approaches zero quickly as the number of users increases. The asymptotic efficiency of the decorrelating detector approaches the line $\eta = 1 - \rho$. At $\rho = 0.5$ and large K, it approaches a 3-dB degradation relative to the optimal detector. This example illustrates the limitations of uncoded detectors. Even with the optimum detector, the asymptotic efficiency degrades when the cross-correlation parameter exceeds 0.5. The K-symmetric channel is not an anomaly in this respect. In fact, the performance of the optimal detector on the K-symmetric channel is an upper bound on the asymptotic efficiency for any uncoded multiuser detector for a bit-synchronous additive Gaussian channel.

D. Asymptotic Efficiency with Coding

For coded systems, the asymptotic efficiency is also given by (10) but only error sequences that correspond to the differences between legitimate codewords are considered in the minimization. In the following we only consider the case where all the channel gains are one $W = I$. Let $\rho_{\max}$ represent the magnitude of the largest off-diagonal element of cross-correlation matrix H.

Theorem 1: With synchronous equal power users, the asymptotic multiuser coded efficiency over a Gaussian channel is upper bounded by

$$\eta_k \leq \min\{1, 2(1 - \rho_{\max})\} \tag{12}$$

when all users apply the same code.

This result indicates that a multiuser system, coded as in Theorem 1, suffers from the same limitations as the uncoded system as the correlation between users increases. A proof is given in the appendix.

Alternatively, we can assume that all users use the same FEC code but that each user has a different pseudorandom interleaver. The analysis of [9] does not apply in this case, as the trellis structure of the decoder is no longer as straightforward and the optimum detector under these conditions is even further from practical. In this case the asymptotic efficiency will depend upon the choice of interleaver. Let $\eta_{k,N}$ be a random variable representing the asymptotic efficiency of the kth user for a set of K independently chosen random interleavers of length N.

Theorem 2 (Optimal Asymptotic Efficiency with Coding and Random Interleaving): For a synchronous Gaussian channel with K equal power users and K randomly selected interleavers of length N

$$\lim_{N\to\infty} \Pr[\eta_{k,N}=1]=1, \qquad k=1,\cdots,K \tag{13}$$

when the cross-correlation matrix H is positive definite.

This theorem is proven in the appendix. It highlights the importance of not separating the detection and decoding problems. In the uncoded case the efficiency of even the optimum detector degrades for $\rho > 0.5$. This loss disappears if the detection process is not separated from the decoding process and one employs random interleaving. The benefits of random interleaving are somewhat analogous to what is observed with turbo codes [15], [16], although, with multiuser decoding, it does not rely on the use of systematic recursive convolutional codes. The positive definiteness of H is a technical requirement of the proof that does not appear necessary in practice. This result is more general than indicated here. In particular, the following corollary comes immediately from the inspection of the proof of the theorem.

Corollary 1: The results of Theorem 2 apply when the cross correlation H is time varying.

It does not matter whether this time variation is intentional or nonintentional. This implies that the results can be applied to cases such as FDMA with a high degree of spectral overlap. In the latter the pulse shapes may be constant from one symbol period to the next, but the effect of frequency offsets and phasing will vary from one symbol period to the next.

III. Multiuser Decoding Algorithm

A. Optimum Algorithm

The following iterative decoding structure can be derived either from intuition or theoretically from a minimum cross-entropy (MCE) framework [17], [24], [25]. Cross-entropy minimization is a statistical inference scheme that estimates the probability distribution function that satisfies given constraints on its moments and minimizes the cross entropy (relative entropy) with respect to an *a priori* distribution [17]. To reduce confusion when applied to the decoding problem, we will use the term *intrinsic* distribution where MCE literature typically uses the term *a priori* distribution. The *intrinsic* distribution represents the distribution $q_o[\mathbf{b}]$ implied by the channel samples $\mathbf{y}$. This is quite different than the usual interpretation of *a priori* information. The constraints correspond to the parity check equations of the code [24], [25]. For example, the moment constraint corresponding to a parity check equation $b_k(1)b_k(2)b_k(3) = 1$ is $E[b_k(1)b_k(2)b_k(3) - 1] = 0$. The general solution to the cross-entropy minimization problem is given by the following lemma [17].

Lemma 1: Let $\mathbf{b}$ be a random vector with *intrinsic* distribution $q_o[\mathbf{b}]$. Let $E[f_i(\mathbf{b})] = 0,\ i = 1,\cdots,F$ be constraints on the moments of $\mathbf{b}$. Then the MCE distribution is given by

$$p[\mathbf{b}] = Aq_o[\mathbf{b}]g_1(\mathbf{b})\cdots g_F(\mathbf{b}) \tag{14}$$

where $g_i(\mathbf{b}) = \exp\{-\lambda_i f_i(\mathbf{b})\}$ for some constants $\{\lambda_i\}$, and the constant A normalizes the probability mass.

For coding applications, the expression for $g_i(\mathbf{b})$ simplifies to $I_i(\mathbf{b})$, that is, the indicator function for the set of codewords that satisfy the parity equation $f_i(\mathbf{b}) = 0$, and the MCE distribution is equivalent to the *a posteriori* codeword distribution [24], [25]. Furthermore, one can iteratively determine the MCE distribution [21], [22]. With this algorithm, the constraints are grouped into sets $C_i(\mathbf{b}),\ i = 1,\cdots,K$ and each of the constraint sets is considered separately. We use the notation $q_o(\mathbf{b})\circ C_i(\mathbf{b})$ to represent the MCE distribution corresponding to the *intrinsic* distribution $q_o[\mathbf{b}]$ and constraint set $C_i(\mathbf{b})$.

Lemma 2: For the *intrinsic* distribution $q_o[\mathbf{b}]$ and constraint sets $C_i(\mathbf{b}), i = 1,\cdots,K$, let $p_1[\mathbf{b}] = q_o[\mathbf{b}]\circ C_1(\mathbf{b})$ and let

$$p_{j+1}[\mathbf{b}] = p_j[\mathbf{b}]\circ C_{1+j \bmod K}(\mathbf{b}) \tag{15}$$

then $p_j[\mathbf{b}]$ converges to the unique MCE distribution with respect to $q_o[\mathbf{b}]$, satisfying all of the constraints.

This is shown in [21, Th. 3.2]. This theorem says that the sequence of local MCE distributions, obtained by applying the constraint sets repeatedly, converges to the global MCE distribution. This forms the basis for the iterative detection/decoding algorithm. In this application the K constraint sets correspond to the interleaver/FEC code combinations for each user. The following two lemmas describe two variations on this algorithm that are useful in practice and retain its optimality.

Lemma 3 (Equivalence of Parallel and Serial Implementations): For the decoding problem, let $p_{\text{serial}}[\mathbf{b}]$ be the result after K steps of the serial algorithm defined by (15). Let $p^k[\mathbf{b}] = q_o[\mathbf{b}]\circ C_k(\mathbf{b}),\ k = 1,\cdots,K$ be the K MCE distributions after one parallel step, then the distribution

$$p_{\text{parallel}}[\mathbf{b}] = c_{\text{parallel}}\frac{p^1[\mathbf{b}]p^2[\mathbf{b}]\cdots p^K[\mathbf{b}]}{(q_o[\mathbf{b}])^{K-1}} = p_{\text{serial}}[\mathbf{b}] \tag{16}$$

where the constant c_{parallel} normalizes the probability mass.

Thus, for the decoding problem, the constraints can be applied in parallel rather than serially. The proof of Lemma 3 is presented in the Appendix, and it follows by induction that the parallel version converges to the same limit as the serial version. The combination of these three lemmas gives the algorithm structure shown in Fig. 3. The algorithm consists of two parts: a combiner and K parallel decoders. Each decoder only considers the bits of one constraint set, i.e., one user. The combiner uses the output of the decoders together with the input to form a new MCE distribution.

The MCE distribution (14) consists of two parts, the *intrinsic* part $q_o[\mathbf{b}]$ and an extrinsic part due to the constraints

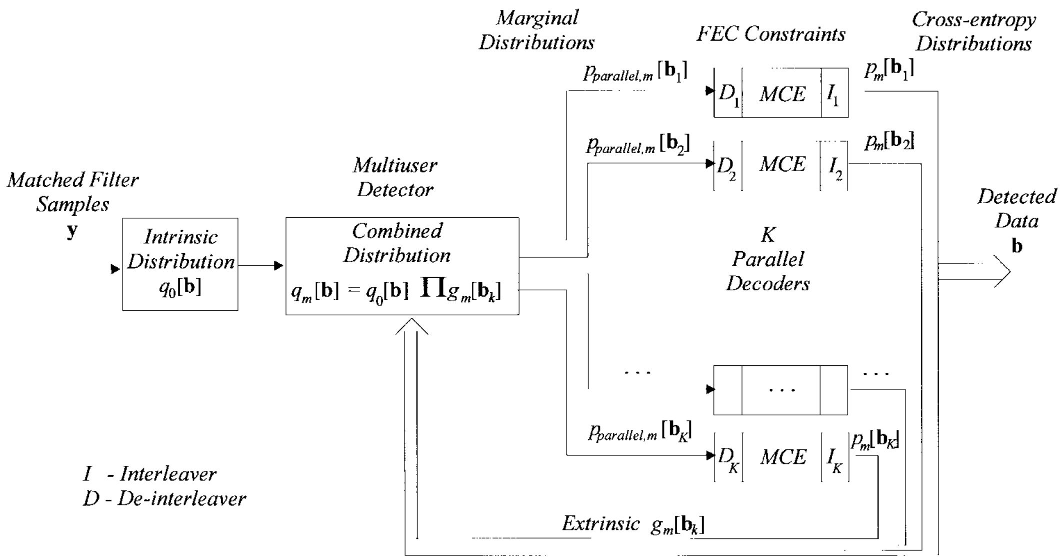

Fig. 3. Illustration of iterative multiuser decoding algorithm.

$g_i(\mathbf{b})$. The nomenclature of *intrinsic* and *extrinsic* comes from an analogy to turbo coding [19]. From Lemma 3

$$p_{\text{parallel}}[\mathbf{b}] = c_{\text{parallel}} q_o[\mathbf{b}] g_1(\mathbf{b}) \cdots g_K(\mathbf{b}) \tag{17}$$

that is, the MCE distribution after one parallel step is the combination of the intrinsic distribution and the extrinsic distribution from each component branch in the step. In Fig. 3 the input to each decoder is the combination of the *intrinsic* distribution obtained from the channel with the *extrinsic* distribution for each user provided by the previous decoding. On the first iteration there is no previous extrinsic distribution.

Lemma 4 (Discarding Extrinsic Data): When the constraint sets concern nonoverlapping subsets of the data $C_k(\mathbf{b}) = C_k(\mathbf{b}_k)$, let

$$p^k_{\text{parallel}}[\mathbf{b}] = c^k_{\text{parallel}} q_o[\mathbf{b}] g_1(\mathbf{b}_1) \cdots g_{k-1}(\mathbf{b}_{k-1}) \cdot g_{k+1}(\mathbf{b}_{k+1}) \cdots g_K(\mathbf{b}_K) \tag{18}$$

then

$$p_{\text{parallel}}[\mathbf{b}] \circ C_k(\mathbf{b}_k) = p^k_{\text{parallel}}[\mathbf{b}] \circ C_k(\mathbf{b}_k). \tag{19}$$

That is, we can ignore previous extrinsic distribution for a subset of bits when calculating a new extrinsic for the same subset of bits. The proof of this lemma is in the appendix. These results can be combined to give Theorem 3, which follows from the observation at the end of Lemma 1 that the global MCE distribution is the *a posteriori* codeword distribution.

Theorem 3 (Optimal Iterative Multiuser Detector): With the algorithm defined by Lemma 2 and the possible modifications defined by Lemmas 3 and 4

$$p_j[\mathbf{b}] \rightarrow \Pr[\mathbf{b} \mid \mathbf{y}] \tag{20}$$

that is, there is convergence to the *a posteriori* codeword distribution.

B. Suboptimum Algorithm

Although optimum, the iterative MCE approach is far from practical. Since it deals with distributions, it requires the determination of the probabilities at the 2^{KN} possible binary vectors. To make it practical requires an additional assumption that *the symbol probabilities at the output of the MCE decoders are independent in distinct symbol intervals*. This assumption is reasonable if the bits of a user codeword are randomly interleaved. This does not assume that bits from different users in the same bit period are independent.

With this independence assumption, the decoders need only output a soft value for each of KN bits rather than 2^{KN} possible codewords. Consequently, instead of the true MCE algorithm, one can use a soft-output decoding algorithm such as [20] for estimating the symbol probabilities. The overall structure of this modified algorithm is best illustrated by an example.

For a memoryless Gaussian channel with matched filter outputs $\{y_k(i)\colon k = 1, \cdots, K,\ i = 1, \cdots, N\}$, the *intrinsic* distribution for the ith symbol period is given by

$$q[\mathbf{b}(i)] = \frac{\det(H)^{-1/2}}{(2\pi\sigma^2)^{K/2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}(i) - H\mathbf{b}(i))^T \cdot H^{-1}(\mathbf{y}(i) - H\mathbf{b}(i))\right\} \tag{21}$$

where the noise has a covariance $\sigma^2 H$. Given the independence assumption, the input to the decoders is, from (17), $p_{\text{parallel},m}[\mathbf{b}(i)]$ for a particular iteration and bit interval where the subscript m implies the distribution after m iterations. Since each decoder acts only on a subset of the bits $\mathbf{b}_k$, the kth decoder input is the corresponding marginal distribution

$$p_{\text{parallel},m}[b_k(i)] = \sum_{b_1(i), \cdots, b_{k-1}(i), b_{k+1}(i), \cdots, b_K(i)} p_{\text{parallel},m}[\mathbf{b}(i)]. \tag{22}$$

Using the independence assumption, the soft-output decoders estimate the symbol probabilities for the ith bit of the

kth user $p_m^k[b_k(i)]$, where

$$p_m^k[b_k(i)] = p_{\text{parallel},m}[b_k(i)]g_{k,m}(b_k(i)). \quad (23)$$

The normalized extrinsic distribution $g'_{k,m}[b_k(i)]$ is then given by

$$g'_{k,m}[b_k(i)=1] = A_{k,m}\frac{p_m^k[b_k(i)=1]}{p_{\text{parallel},m}[b_k(i)=1]} \quad (24)$$

for $b_k(i) = 1$ and similarly for $b_k(i) = -1$, where $A_{k,m}$ normalizes the sum of the probabilities of "-1" and "1." The combined intrinsic and extrinsic distribution is given by

$$p_{\text{parallel},m+1}[\mathbf{b}(i)] = A_m q_o[\mathbf{b}(i)]g'_{m,1}[b_1(i)]\cdots g'_{m,K}[b_K(i)] \quad (25)$$

where the constant A_m normalizes the probability mass. At any time, a decision can be made based on the decoder output. In practice, better performance is obtained if one uses $p^k_{\text{parallel},m}[\mathbf{b}(i)]$, i.e., the mth iteration of $p^k_{\text{parallel}}[\mathbf{b}(i)]$ as per Lemma 4, instead of $p_{\text{parallel},m}[\mathbf{b}(i)]$ in the right-hand side of (22). Ostensibly, this is because it increases the validity of the independence assumption. This provides a manageable algorithm that is suitable for multiuser detection. From a complexity viewpoint, there are K decoders (or a single decoder reused K times) that is matched to the code of a single user. The complexity of this algorithm can be expressed as $(O(2^K) + O(2^\kappa))$ per bit per iteration where κ is the code constraint length.

IV. Simulation Results

In this section the simulation results for the proposed iterative detection algorithm are presented for five and ten users with a variety of correlation values ρ. All simulations use a block size of 500 information bits for each user. Each user uses the same rate-1/2 constraint length 5 convolutional code with generators [10 011] and [11 101]. Each user uses a different pseudorandom interleaver, and the same set of interleavers is used for all simulation runs. The interleavers were chosen at random and no attempt was made to optimize them. Each simulation point is tested for the minimum of 1600 errors or four million bits.

A. Five-User Results

In Fig. 4 the performance of the iterative detector on a K-symmetric channel with five users and a ρ of 0.60 is shown for 1, 2, 4, 8, and 16 iterations. Even with this high correlation, performance converges rapidly to single-user performance $(\rho = 0)$ at higher E_b/N_0 ratios as predicted by theory. There is a noticeable loss at low SNR's. In Fig. 5 the performance with five users and a ρ of 0.75 is shown. In this example a threshold effect is becoming apparent at the lower E_b/N_0 ratios. Below the threshold, the channel is unusable. Above this threshold, performance rapidly converges to single-user performance.

In Fig. 6 the performance on a K-symmetric channel with a ρ of 0.90 and five users is shown. In this case the threshold effect is even more apparent, and performance has moved significantly away from the single-user performance curve.

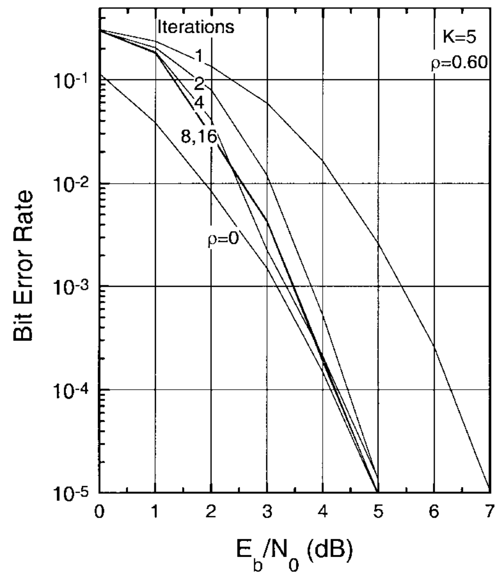

Fig. 4. Comparison of BER performance of iterative decoder with five users and $\rho = 0.60$ (1, 2, 4, 8, and 16 iterations) to ideal single-user performance.

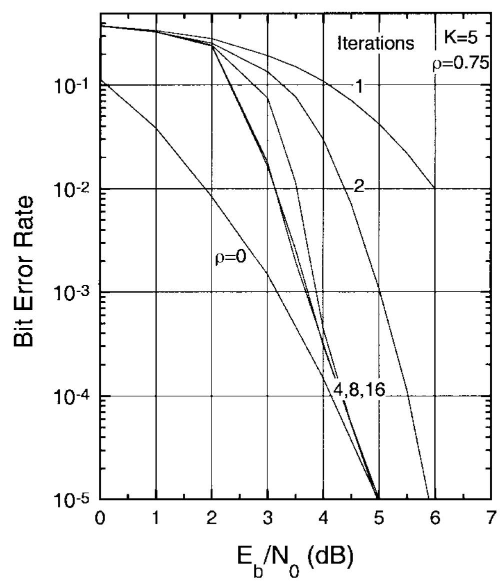

Fig. 5. Comparison of BER performance of iterative decoder with five users and $\rho = 0.75$ (1, 2, 4, 8, and 16 iterations) to ideal single-user performance.

The degradation is 1.5 dB at a BER of 10^{-5}, and further iteration does not improve upon this.

B. Ten-User Results

For comparison purposes, the corresponding performance with ten users and a ρ of 0.60 on a K-symmetric channel is shown in Fig. 7. Like previous cases, there is a threshold E_b/N_0 below which the channel is not usable. Relative to the five-user case with the same correlation, performance on the first iteration is significantly degraded. However, with enough iterations, performance still converges to the single-user performance for E_b/N_0 ratios of 4 dB and higher. It also performs significantly better than the five-user case with a ρ of 0.90. Note that the variance of the multiple-access interference

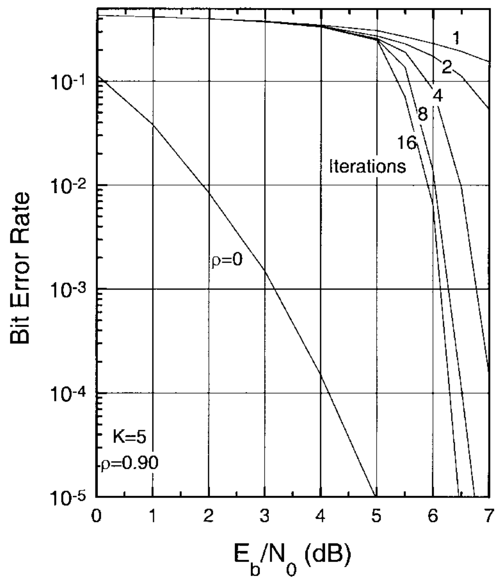

Fig. 6. Comparison of BER performance of iterative decoder with five users and $\rho = 0.90$ (1, 2, 4, 8, and 16 iterations) to ideal single-user performance.

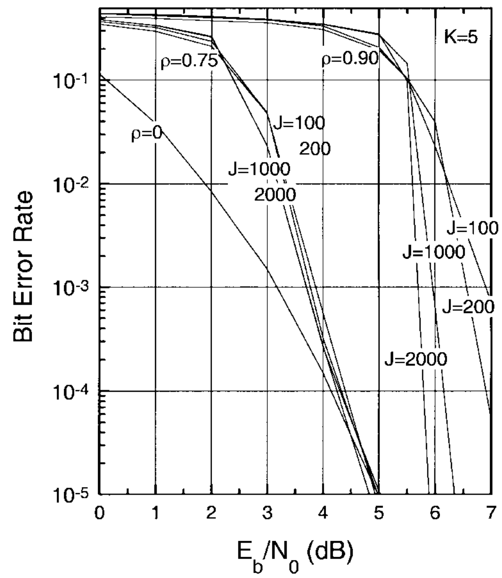

Fig. 8. Dependence of performance on interleaver size for cross correlations of $\rho = 0.75$ and 0.90. Results are for five users, rate-1/2 $k = 5$ convolutional code, and 16 iterations.

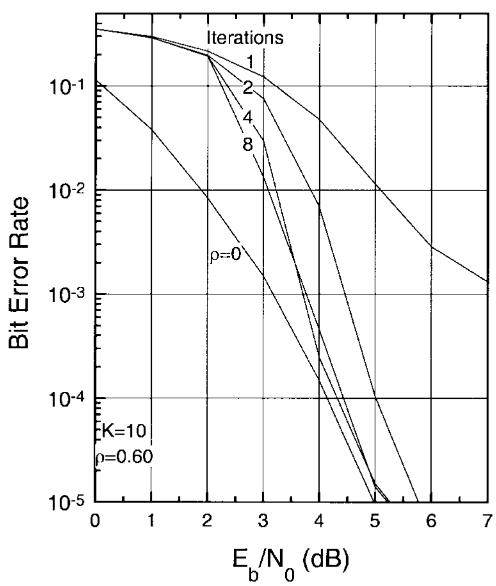

Fig. 7. Comparison of BER performance of iterative decoder with ten users and $\rho = 0.60$ (1, 2, 4, and 8 iterations) to ideal single-user performance.

$\rho^2(K-1)$ is identical in these two cases. The reason why five users $(\rho = 0.9)$ significantly degrade the detector performance while ten users $(\rho = 0.6)$ do not is explained in Section V.

C. Interleaver Dependence

The interleaver used in these simulations was chosen at random. The same interleaver was used for all simulations having the same block size. No attempt was made to optimize this interleaver, as theory indicates that an interleaver picked at random is expected to be good. To test this assumption, we repeated the tests with a number of different interleavers, all of which were chosen at random. The simulated performance results with the different interleavers were virtually identical. Theory also indicates that there will be bad interleavers. For example, with no interleaving, theory claims that performance should degrade as soon as the correlation parameter exceeds 0.5. Simulation confirmed that the expected degradation was observed.

As the analysis indicated, for sufficiently large N, asymptotic multiuser performance is expected to be the same as single-user performance. In Fig. 8 the performance with $K = 5$ users and interleaver sizes corresponding to $J = 100$, 200, 1000, and 2000 information bits $(N = 2(J + \kappa))$ for the constraint length 5 rate-1/2 convolutional code is shown. The results correspond to the K-symmetric channel with cross-correlation values of $\rho = 0.75$ and 0.90 and 16 iterations in each case. For the $\rho = 0.75$ case, the predicted performance is obtained with J as small as 100. For the $\rho = 0.90$ case, there appears to be some degradation with the smaller interleavers although the simulation results are not conclusive. The interleaver size clearly plays a role at low SNR. Larger interleaver sizes improve performance in the range between threshold and single-user performance. The interleaver size has little effect on the threshold behavior of the detector. Although not shown, simulation results indicate that increasing the interleaver size has little effect on the rate of convergence.

D. Code Dependence

The dependence of performance on the FEC is also of interest. For comparison with the constraint length 5 rate-1/2 convolutional code used to this point, rate-1/2 convolutional codes with constraint lengths of 4 and 7 were tested on the K-symmetric channel. A recursive systematic code was also tested. The results are qualitatively similar to the constraint length 5 code in all cases. In particular, performance converged rapidly to single-user performance at higher E_b/N_0 ratios. Furthermore, the threshold was found to be independent of the FEC code used and found to depend only on the number of users K and correlation ρ.

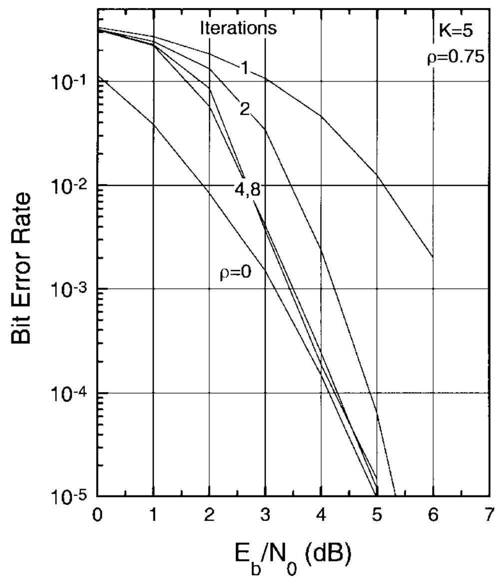

Fig. 9. Comparison of BER performance of iterative decoder with five users and $\rho = 0.75$ (1, 2, 4, and 8 iterations) to ideal single-user performance over the K-exponential channel.

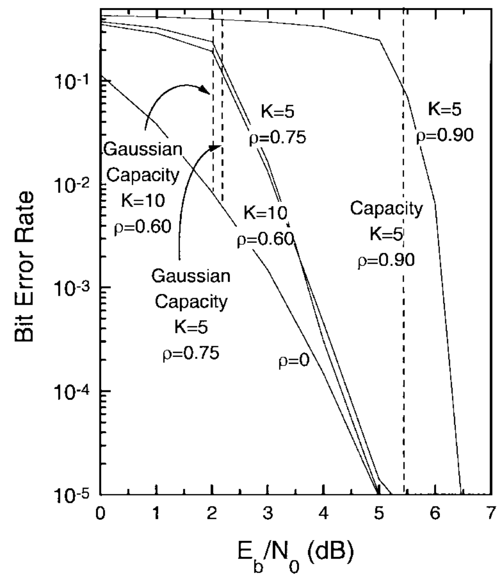

Fig. 10. Comparison of BER performance of various scenarios with theoretical capacity limits of the K-symmetric channel and the single-user bound; all results are after 16 iterations and $J = 500$ information bits.

E. Other Channels

Simulation results, thus far, have concentrated on the K-symmetric channel because it is analytically tractable and allows a comparison of different channels through the single cross-correlation parameter ρ. The theoretical results and algorithms derived here, however, apply to general K-user Gaussian bit-synchronous channels. For example, let the K-*exponential* channel be defined as having the cross-correlation matrix given by

$$H = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{K-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{K-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{K-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{K-2} & \cdots & \cdots & \rho & 1 & \rho \\ \rho^{K-1} & \cdots & \cdots & \rho^2 & \rho & 1 \end{bmatrix}. \quad (26)$$

It has the same asymptotic properties in ρ as the K-symmetric channel; however, for intermediate values of ρ, it is substantially different from the K-symmetric channel. In Fig. 9 the performance obtained over the K-exponential channel for five users with a cross-correlation parameter ρ of 0.75 is shown.

The conclusion is that while quantitatively different, the BER performance is qualitatively the same on the K-exponential channel as observed on the K-symmetric channel. The K-exponential channel is not as harsh as the K-symmetric channel and for this reason there is faster convergence to single-user performance when $\rho = 0.75$. At the higher cross-correlation value of 0.90, the K-exponential channel has a threshold, but at significantly lower E_b/N_0 than observed with the K-symmetric channel.

V. Analysis of Simulation Results

The simulation results have shown the near-optimum performance of the MCE-based decoding algorithm and emphasize that a practical multiuser decoder can be implemented that has minimal losses, if any, with respect to single-user performance in many situations.

The results also show a threshold-type behavior as the correlation parameter ρ increases on the K-symmetric channel. Above the threshold E_b/N_0, performance rapidly approaches single-user performance. Below the threshold, the channel is practically unusable. The fact that this threshold does not depend on either the code or the interleaver length suggests that it may be capacity related. For positive definite H, the discrete-time model of (5) is equivalent to a K-parallel channel with dependent noise [23], and we have the following result.

Theorem 4: The theoretical capacity of the system $\mathbf{y}(i) = H\mathbf{b}(i) + \mathbf{n}(i)$ with independent equal power users and Gaussian signaling is

$$C(\gamma) = \frac{1}{2}\log_2\left(\frac{\det(\gamma I + H^{-1})}{\det(H^{-1})}\right) \text{ bits/dimension} \quad (27)$$

where γ is the single-user SNR per dimension.

For a rate-1/2 code, $\gamma = E_b/N_0$. This theorem can be proven using a simplified version of that given in [23]. For symmetric users, the capacity per user is $1/K$th of this quantity. In Fig. 10 the performance obtained in various K-symmetric scenarios is compared to the combination of the single-user capacity limit, i.e., the E_b/N_0 for a capacity of 1/2 bits/dimension/user, and the single-user BER limit. For two of the three scenarios, $K = 5$ and $\rho = 0.75$, and $K = 10$ and $\rho = 0.60$, performance is within 1 dB of the limits imposed by these two curves. In the third case, it was found that the required E_b/N_0 to achieve 1/2 bits/dimension/user was significantly larger with binary signaling than with Gaussian signaling.[3] In Fig. 10 the capacity limit for the $(K = 5, \rho = 0.9)$ case was

[3]This was pointed out by an anonymous reviewer who noted that the sum-capacity of the binary adder channel $(K = 5, \rho = 1)$ is only 2.2 bits/dimension.

computed numerically for the true signaling distribution, not Gaussian signaling. Performance is within 1 dB of capacity over the illustrated range. With the larger block sizes shown in Fig. 8, the difference is less than 0.5 dB. Similar behavior was found on other channels with ρ values very close to one, as long as the channel capacity was not exceeded.

The K-symmetric channel is an artificial construct developed to test and compare the performance of the detection/decoding algorithm. In many circumstances it will not have practical relevance. The K-exponential channel model will be a closer model of many practical situations. However, the K-symmetric channel does provide meaningful tests of the behavior and robustness of the algorithm and the theory.

VI. Conclusion

In this paper it has been implicitly shown that, in a multiuser system with low user correlations, techniques which perform detection followed by decoding should be capable of near-optimum performance. When the user correlations exceed 0.5, these approaches have limitations. These limitations can be remedied by FEC coding, random interleaving, and joint detection/decoding. Theoretical analysis shows that, with this approach, multiuser detection/decoding is capable of asymptotically achieving optimal single-user performance for user correlations approaching one.

The paper also derives an optimal iterative algorithm for coded multiuser detection/decoding based on iterative techniques for minimizing cross entropy. It is shown that, with an independence assumption, this optimal algorithm becomes a practical algorithm for multiuser detection. Simulation results are presented that validate the theoretical claims. These simulation results also show that, in many scenarios, performance with the practical algorithm is limited only by single-user performance and theoretical capacity limits when using only simple codes and short block lengths.

The complexity of this algorithm can be expressed approximately as $(O(2^K) + O(2^\kappa))$ per bit per iteration where κ is the code constraint length. However, as the simulation results show, one can approach channel capacity with quite small K, κ, and m. In addition, this algorithm for synchronous users can be extended to provide similar performance for asynchronous users with the penalty of increased complexity

Appendix

Proof of Theorem 1: Since we assume $W = I$, let the numerator of the asymptotic efficiency expression (10) be

$$d^2 = \min_{\varepsilon \neq 0} \sum_{n=1}^{N} \varepsilon(n)^T H \varepsilon(n). \tag{28}$$

For a code of minimum distance d_{free}, a single-user error event of minimum distance implies that

$$d^2 \leq d_{\text{free}}. \tag{29}$$

This upper bound can be tightened by noting that, when each of two users have coinciding minimum distance error events, then

$$\begin{aligned} d^2 &= \min_{\varepsilon \neq 0} \sum_{n=1}^{N} [\varepsilon_1(n) \quad \varepsilon_2(n)] \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1(n) \\ \varepsilon_2(n) \end{bmatrix} \\ &= \min_{\varepsilon \neq 0} \sum_{n=1}^{N} (\varepsilon_1(n)^2 + \varepsilon_2(n)^2 + 2\rho\varepsilon_1(n)\varepsilon_2(n)) \\ &\leq 2d_{\text{free}}(1 - |\rho|) \end{aligned} \tag{30}$$

where ρ is the correlation between users and the minimum is achieved with $\varepsilon_1(n) = -\text{sign}(\rho)\varepsilon_2(n)$. With $\rho_{\max}$ as the magnitude of the maximum off-diagonal element of H, substituting (29) and (30) into (10), one has

$$\eta = \frac{d^2}{d_{\text{free}}} \leq \min\{1, 2(1 - \rho_{\max})\}. \tag{31}$$

□

Proof of Theorem 2: In the synchronous Gaussian channel the asymptotic efficiency is given by (10). For the case $W = I$, we have

$$\begin{aligned} \Pr[\eta_{k,N} < 1] &= \Pr\left[\min_{\varepsilon_k \neq 0} \delta^2(\varepsilon) < d_{\text{free}}\right] \\ &\leq \Pr\left[\min_{\varepsilon \neq 0} \delta^2(\varepsilon) < d_{\text{free}}\right] \end{aligned} \tag{32}$$

where

$$\delta^2(\varepsilon) = \sum_{n=1}^{N} \varepsilon(n)^T H \varepsilon(n). \tag{33}$$

Define the weight of a multiuser error event $|\varepsilon|$ as the number of nonzero elements $\varepsilon_k(n)$ in the sequence of vectors $\{\varepsilon(n), n = 1, \cdots, N\}$. For a particular error event, let K^* be the number of users contributing to that error event. For $K^* = 1$, it is clear that $\delta^2(\varepsilon) \geq d_{\text{free}}$. Let $\lambda_{\min}$ be the minimum eigenvalue of H and let $D = d_{\text{free}}/\lambda_{\min}$. Then, from (33), for $|\varepsilon| > D$, $\delta^2(\varepsilon) > d_{\text{free}}$. Thus, only $K^* \geq 2$ and error events of weight between $2d_{\text{free}}$ and D (inclusive) must be considered. Then, from the union bound

$$\Pr[\eta_{k,N} < 1] \leq \sum_{d=2d_{\text{free}}}^{D} A_d \max_{|\varepsilon|=d} \Pr[\delta^2(\varepsilon) < d_{\text{free}}] \tag{34}$$

where A_d is the number of error events of weight d. Assume a common convolutional code for all users with transfer function

$$c(Z) = \sum_{d \geq d_{\text{free}}} c_d Z^d. \tag{35}$$

Let $c_{\max} = \max\{c_d\colon d_{\text{free}} \leq d \leq D\}$. Then, for a given K^* and any interleaver of length N, the number of error events of weight d, where $2d_{\text{free}} < d < D$, is upper bounded by

$$A_d \leq (c_{\max}N)^{K^*} \leq (c_{\max}N)^{d/d_{\text{free}}} \tag{36}$$

since $d \geq K^* d_{\text{free}}$. From the following Lemma 5, we have that

$$\max_{|\varepsilon|=d} \Pr[\delta^2(\varepsilon) < d_{\text{free}}] \leq c_{\text{bin}} \left(\frac{d}{N}\right)^{d-d_{\text{free}}+1}. \tag{37}$$

Substituting (36) and (37) into (34), one obtains

$$\Pr[\eta_{k,N} < 1] \le \sum_{d=2d_{\rm free}}^{D} (c_{\max} N)^{d/d_{\rm free}} c_{\rm bin} \left(\frac{d}{N}\right)^{d-d_{\rm free}+1} \le c_{\rm sum} N^{\mu} \quad (38)$$

where $c_{\rm sum} = c_{\rm bin} c_{\max}^{D/d_{\rm free}} D^{D-d_{\rm free}+2}$ is determined by upper bounding d by D, and

$$\mu = \max_{2d_{\rm free} \le d \le D} \{d/d_{\rm free} - d + d_{\rm free} - 1\} = 1 - d_{\rm free} \quad (39)$$

by letting $d = 2d_{\rm free}$. Thus, for $\lambda_{\min} > 0$ and $d_{\rm free} \ge 2$, $\lim_{N\to\infty} \Pr[\eta_{k,N} < 1] = 0$. □

Proof of Lemma 3: In [24] and [25] it is shown that, for the decoding problem, the MCE distribution is given by

$$p[x] \circ C_i(x) = A_i p(x) I_{C_i}(x) \quad (40)$$

where A_i is a constant that normalizes the probability mass and $I_C(x)$ is the indicator function for the set C, i.e., codewords satisfying the parity check equations in the constraint set C.

Consequently,

$$p_{\rm serial}[\mathbf{b}] = A_{\rm serial} q_o[\mathbf{b}] I_{C_1}[\mathbf{b}] I_{C_2}(\mathbf{b}), \cdots, I_{C_K}(\mathbf{b}) \quad (41)$$

and using (40) in (16), one obtains

$$p_{\rm parallel}[\mathbf{b}] = A_{\rm parallel} q_o[\mathbf{b}] I_{C_1}(\mathbf{b}), \cdots, I_{C_K}(\mathbf{b}) \quad (42)$$

and, consequently, $p_{\rm serial}[\mathbf{b}] = p_{\rm parallel}[\mathbf{b}]$. □

Proof of Lemma 4: The distributions $p_{\rm parallel}[\mathbf{b}]$ and $p^k_{\rm parallel}[\mathbf{b}]$ are defined by (42) and (18), respectively. Then, using (40) and since $(I_{C_k}(\mathbf{b}))^2 = I_{C_k}(\mathbf{b})$, it follows immediately that

$$p_{\rm parallel}[\mathbf{b}] \circ C_k(\mathbf{b}) = p^k_{\rm parallel}[\mathbf{b}] \circ C_k(\mathbf{b}). \quad (43)$$

This illustrates the interesting point that when the true MCE algorithm is applied to the decoding problem, convergence occurs in one cycle through the constraint sets. It is only when one uses approximations to the true MCE algorithm that more than one iteration will bring potential improvement. □

Lemma 5: For an error event over $K^* \ge 2$ users

$$\max_{|\varepsilon|=d} \Pr[\delta^2(\varepsilon) < d_{\rm free}] \le c_{\rm bin} \left(\frac{d}{N}\right)^{d-d_{\rm free}+1}. \quad (44)$$

Proof: In any symbol interval an error can be a singleton (due to a single user) or otherwise (due to coincidence of multiple user symbol errors). Let s_d be a random variable representing the number of singletons in an error event of weight d; this depends upon the interleaver and the error event. Then the corresponding distance is lower bounded by

$$\delta^2(\varepsilon) = \sum_{n=1}^{N} \varepsilon(n)^T H \varepsilon(n) \ge \sum_{n:|\varepsilon(n)|=1} \varepsilon(n)^T H \varepsilon(n) = s_d \quad (45)$$

where ε is any error event of weight d and $|\varepsilon(n)|$ is the number of user errors in symbol n. For any ordering of the individual user errors, let X_i be a binary random variable (0 or 1) representing whether the ith error is a singleton or not. Then

$$s_d = \sum_{i=1}^{d} X_i. \quad (46)$$

Using the chain rule for the joint cumulative distribution $F(X_1, X_2, \cdots, X_d)$ one has

$$F(X_1, \cdots, X_d) = F_1(X_1) F_2(X_2 \mid X_1) \cdots F_d(X_d \mid X_1, \cdots, X_{d-1}) \le B_p(X_1) B_p(X_2) \cdots B_p(X_d) \quad (47)$$

where, for a random interleaver, the individual distributions in the product $F_i(\ )$ have been upper bounded by the Bernoulli distribution with parameter $p = (1 - d/N)$. For general cumulative distribution functions F_i and B_i, if $F_i(X) \le B_i(X)$ and $F_j(X) \le B_j(X)$, then $F_i(X) * F_j(X) = \int F_i(X - s)\, dF_j(s) \le B_i(X) * B_j(X)$. Consequently, from the bound in (47), the distribution of s_d is upper bounded by the binomial distribution of d trials with parameter p

$$\begin{aligned}\Pr[s_d < d_{\rm free}] &\le \sum_{j=0}^{d_{\rm free}-1} \binom{d}{j} (1-p)^{d-j} p^j \\ &\le (1-p)^{d-d_{\rm free}+1} \sum_{j=0}^{d_{\rm free}-1} \binom{d}{j} p^j \\ &\le (1-p)^{d-d_{\rm free}+1} \sum_{j=0}^{d_{\rm free}-1} \frac{(dp)^j}{j!} \\ &< (1-p)^{d-d_{\rm free}+1} e^D p \end{aligned} \quad (48)$$

and (44) follows for a constant $c_{\rm bin} < e^D p$. □

ACKNOWLEDGMENT

The author would like to thank the Editor, S. S. Pietrobon, and the anonymous reviewers for a very thorough review and many helpful suggestions.

REFERENCES

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

[2] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

[3] ——, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, pp. 496–508, Apr. 1990.

[4] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.

[5] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.

[6] A. Duel-Hallen, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels," *IEEE Trans. Commun.*, vol. 43, pp. 421–434, Feb./Mar./Apr. 1995.

[7] 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, pp. 641–649, May 1990.

[8] P. Hoeher, "On channel coding and multiuser detection for DS-CDMA," in *Proc. 2nd Int. Conf. Universal Personal Communications*, Ottawa, Ont., Canada, Oct. 1993, pp. 641–646.

[9] T. R. Giallorenzi and S. G. Wilson, "Multiuser ML sequence estimator for convolutional coded asynchronous DS-CDMA systems," *IEEE Trans. Commun.*, vol. 44, pp. 997–1008, Aug. 1996.
[10] A. Hafeez and W. E. Stark, "Combined decision-feedback multiuser detection/soft-decision decoding for CDMA channels," in *Proc. Vehicular Technology Conf.*, Atlanta, GA, Apr. 1996.
[11] R. Jana and L. Wei, "Bounds for minimum Euclidean distance for coded multiuser CDMA systems," in *Proc. 4th Int. Symp. Spread Spectrum Techniques*, Mainz, Germany, 1996, pp. 847–851.
[12] A. Grant and C. Schlegel, "An iterative multiuser receiver: The consensus detector," in *Proc. ISIT'95*, Whistler, Canada, Sept. 1995, p. 383.
[13] T. R. Giallorenzi and S. G. Wilson, "Suboptimum multiuser receivers for convolutionally coded asynchronous DS-CDMA systems," *IEEE Trans. Commun.*, vol. 44, pp. 1183–1196, Sept. 1996.
[14] M. Moher, "Cross-entropy and iterative detection," Ph.D. dissertation, Carleton Univ., Ottawa, Ont., Canada, May 1997.
[15] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correction coding and decoding: Turbo codes," in *Proc. ICC*, Geneva, Switzerland, May 1993, pp. 1064–1070.
[16] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 409–428, Mar. 1996.
[17] J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 472–482, July 1981.
[18] M. C. Reed *et al.*, "Iterative multi-user detection for DS-CDMA with FEC," in *Proc. Int. Symp. Turbo Codes*, Brest, France, Sept. 1997, pp. 162–165; also in *Proc. IEEE PIMRC*, Helsinki, Finland, Sept. 1997, pp. 740–744.
[19] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 429–445, Mar. 1996.
[20] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.
[21] I. Csiszar, "I-divergence geometry of probability distributions and minimization problems," *Ann. Prob.*, vol. 3, pp. 146–158, 1975.
[22] C. T. Ireland and S. Kullback, "Contingency tables with given marginals," *Biometrica*, vol. 55, pp. 179–188, 1968.
[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[24] M. Moher and A. Gulliver, "Cross-entropy and iterative detection," *IEEE Trans. Inform. Theory*, to be published.
[25] M. Moher, "Decoding via cross-entropy minimization," in *Proc. IEEE GLOBECOM*, Houston, TX, 1993, pp. 809–813.
[26] B. R. Vojcic, Y. Shama, and R. L. Pickholtz, "Optimum soft output MAP detector for coded multiuser communications," in *Proc. Int. Symp. Information Theory*, June–July 1997, p. 229.

Michael L. Moher (M'84) received the B.A.Sc. degree in engineering science and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, Ont., Canada, and the Ph.D. degree in systems and computer engineering from Carleton University, Ottawa, Ont., Canada.

From 1980 to 1982 he was with Telesat Canada, where he was responsible for the analysis of new and upgraded thin route satellite links and was involved in the field trials of a new 90 Mb/s *Ku*-band data service. From 1982 to 1988 he led various projects and studies at Miller Communications, Kanata, that analyzed, simulated, and implemented hardware prototypes of advanced communications systems for military, broadcast, and mobile communications. Since 1988 he has been with the Communications Research Centre, Ottawa, Ont., Canada, where he has led various proof-of-concept modem developments for mobile satellite applications, participated in international standards activities, and acted as an industrial consultant for several new wireless and mobile satellite systems. His current research interests are in the area of digital signal processing and detection, and multiuser decoding.