

Iterative Multiuser Joint Decoding: Optimal Power Allocation and Low-Complexity Implementation

Giuseppe Caire, *Member, IEEE*, Ralf R. Müller, *Member, IEEE*, and Toshiyuki Tanaka, *Member, IEEE*

Abstract—We consider a canonical model for coded code-division multiple access (CDMA) with random spreading, where the receiver makes use of iterative *belief-propagation* (BP) joint decoding. We provide simple *density-evolution* analysis in the large-system limit (large number of users) of the performance of the BP decoder and of some suboptimal approximations based on *interference cancellation* (IC). Based on this analysis, we optimize the received user signal-to-noise ratio (SNR) distribution in order to maximize the system spectral efficiency for given user channel codes, channel load (users per chip), and target user bit-error rate (BER). The optimization of the received SNR distribution is obtained by solving a simple linear program and can be easily incorporated into practical *power control* algorithms. Remarkably, under the optimized SNR assignment, the suboptimal minimum mean-square error (MMSE) IC-based decoder performs almost as well as the more complex BP decoder. Moreover, for a large class of commonly used convolutional codes, we observe that the optimized SNR distribution consists of a finite number of discrete SNR levels. Based on this observation, we provide a low-complexity approximation of the MMSE-IC decoder that suffers from very small performance degradation while attaining considerable savings in complexity. As by-products of this work, we obtain a closed-form expression of the multiuser efficiency (ME) of power-mismatched MMSE filters in the large-system limit, and we extend the analysis of the symbol-by-symbol maximum *a posteriori* probability (MAP) multiuser detector in the large-system limit to the case of nonconstant user powers and nonuniform symbol prior probabilities.

Index Terms—Iterative decoding, multiple-access channel capacity, multiuser detection, statistical mechanics.

I. PROBLEM STATEMENT AND PRIOR WORK

THE *canonical* real-valued model for the Gaussian multiple-access discrete-time waveform channel is given by [1]

$$\mathbf{y}_n = \mathbf{S}\mathbf{W}\mathbf{x}_n + \boldsymbol{\nu}_n, \quad n = 1, \dots, N \quad (1)$$

Manuscript received March 12, 2003; revised February 6, 2004. The work of T. Tanaka was supported by EPSRC under Research Grant GR/NR00562 and Grant-in-Aid for scientific research priority areas 14084209, MEXT, Japan. The work of R. R. Müller was supported by the Kplus Program of the Austrian Government. The material in this paper was presented in part at the 39th Allerton Conference on Communication, Control and Computing, Monticello, IL, October 2001, at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002, and at the 40th Allerton Conference on Communication, Control and Computing, Monticello, IL, October 2002.

G. Caire is with the Institut Eurecom, B.P. 193, 06904 Sophia-Antipolis, France (e-mail: caire@eurecom.fr).

R. Müller is with the Forschungszentrum Telekommunikation Wien FTW, Tech Gate Vienna, 1220 Vienna, Austria (e-mail: mueller@ftw.at).

T. Tanaka is with the Tokyo Metropolitan University, Tokyo 192-0397, Japan (e-mail: tanaka@eei.metro-u.ac.jp).

Communicated by V. V. Veeravalli, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2004.833351

where $\mathbf{x}_n \in \mathbb{R}^K$, $\mathbf{y}_n, \boldsymbol{\nu}_n \in \mathbb{R}^L$ are the input, output, and noise signal vectors at time n , respectively, $\mathbf{S} \in \mathbb{R}^{L \times K}$ is a matrix containing by columns the user discrete-time signature waveforms (spreading sequences) \mathbf{s}_k , of length L samples, and $\mathbf{W} = \text{diag}(w_1, \dots, w_K)$ contains the user amplitudes. The noise is Gaussian independent and identically distributed (i.i.d.), with variance per component σ_0^2 (we write $\boldsymbol{\nu}_n \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I})$).

We let \mathcal{C}_k denote the user codebooks, of rate $R_k = \frac{1}{N} \log_2 |\mathcal{C}_k|$ bit per symbol. As usual, in multiple-access channels the users send independent and independently encoded information [2]. Let $x_{k,n}$ denote the n th coded symbol of user k . We assume that $\mathbb{E}[x_{k,n}] = 0$, $\mathbb{E}[x_{k,n}^2] = 1$, and $\mathbb{E}[x_{k,n}x_{j,i}] = 0$ for $(j,i) \neq (k,n)$. Moreover, we normalize the user signature waveforms such that $\|\mathbf{s}_k\|^2 = 1$, so that $\gamma_k \triangleq w_k^2/\sigma_0^2$ takes on the meaning of *received* signal-to-noise ratio (SNR) for user k .

Each k th user, in order to transmit its information message $m_k \in \{1, \dots, |\mathcal{C}_k|\}$, sends the codeword

$$\phi_k(m_k) = (x_{k,1}, \dots, x_{k,N}) \in \mathcal{C}_k$$

in N consecutive channel uses as given in (1). At the receiver, a *joint decoder* maps the received signal $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ into a K -tuple of information messages $(\hat{m}_1, \dots, \hat{m}_K)$. Without loss of generality, we assume that the user information messages are represented by vectors of B_k *information bits* \mathbf{b}_k (e.g., \mathbf{b}_k can be seen as the binary representation of the index m_k). Hence, we define the per-user bit-error rate (BER) as

$$P_b^{(k)} = \frac{1}{B_k} \sum_{j=1}^{B_k} \Pr(\hat{b}_{k,j} \neq b_{k,j}) \quad (2)$$

under the usual assumption that the user information messages are uniformly distributed.

From standard arguments [3, Ch. 8]), we have that the transmitted signal bandwidth is L/T_s , where T_s is the (continuous-time) duration of one channel use. Therefore, the system spectral efficiency is given by [4]

$$\rho = \frac{1}{L} \sum_{k=1}^K R_k \quad \text{bit/s/Hz.} \quad (3)$$

We shall also define the *system* received energy-per-bit

$$E_b \triangleq \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K R_k}$$

and the system E_b/N_0 , given by [4], [5]

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} \triangleq \frac{E_b}{2\sigma_0^2} = \frac{1}{L} \sum_{k=1}^K \gamma_k. \quad (4)$$

The model (1) has been used extensively in order to derive in simple and concise form most *multiuser detection* (MUD) algorithms (see [1] and references therein) and several recent results on the performance analysis of MUD algorithms in the large-system limit (i.e., letting both K and L go to infinity with fixed ratio $K/L = \alpha$) under the *random spreading* assumption, i.e., letting the entries of \mathbf{S} be generated i.i.d. according to some probability distribution (see, for example, [4], [6]–[13]).

In this work, we are concerned with the practically relevant problem of maximizing the system spectral efficiency ρ for a given family of user codes $\{\mathcal{C}_k : k = 1, \dots, K\}$, given iterative joint decoders (see [14] and references therein) and subject to the individual maximum BER constraints $P_b^{(k)} \leq \epsilon$ for all $k = 1, \dots, K$, under the random-spreading assumption and in the large-system limit. We conclude this section by reviewing some known results on spectral efficiency of random-spreading code-division multiple access (CDMA) and by providing a preview of the remainder of this paper.

Maximum spectral efficiency with optimal coding/decoding and vanishing BER. The maximum spectral efficiency of random-spreading CDMA with no restrictions on coding and decoding and for vanishing BER (i.e., $\epsilon \rightarrow 0$)¹ was found by Verdú and Shamai in [4], [5] for given finite *channel load* $K/L = \alpha$, and reads

$$C = \frac{\alpha}{2} \log_2(1 + \gamma\eta) - \frac{1}{2} \log_2 \eta - \frac{1 - \eta}{2} \log_2 e \quad \text{bit/s/Hz} \quad (5)$$

where η is the solution to [6]

$$\frac{1}{\eta} = 1 + \alpha \frac{\gamma}{1 + \eta\gamma}. \quad (6)$$

The optimal $(E_b/N_0)_{\text{sys}}$ for given α and C is given by

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{\alpha\gamma}{2C}. \quad (7)$$

The spectral efficiency (5) is achieved by Gaussian codebooks and constant user received SNR. We say that the received power distribution is *constant* if all users are received at the same SNR level γ , i.e., the empirical cumulative distribution function of the received SNRs is a unit step with jump at γ .

The supremum of C over $\alpha \geq 0$ is obtained for $\alpha \rightarrow \infty$ (an infinite number of users per dimension, with vanishing user coding rate), and is given by the single-user Gaussian channel spectral efficiency

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{2^{2C} - 1}{2C}. \quad (7)$$

It is interesting to notice that, in order to achieve (5), an optimal (maximum-likelihood (ML)) joint decoder is not necessary. In fact, the same optimal spectral efficiency is achieved by a *stripping* decoder that considers the users in sequence (say, in the order $k = 1, 2, \dots, K$) and, at each stage k , decodes the k th message based on the linear MMSE estimate of the k th user codeword from the received signal after subtracting the already decoded users [15]. The price incurred by stripping is that the user coding rates must be assigned such that the transmitted

rate K -tuple coincides with a successively decodable point of the multiple-access capacity region [2] or, if equal user rates are desired, the user received SNRs must be assigned such that the equal-rate point is successively decodable (at the price of some loss in the total achievable rate). The power/rate assignment with practical families of user codes (notably, low-density parity-check (LDPC) codes) for successive stripping decoding is studied in [16].

The spectral efficiency with optimum joint decoding in the case of constant received SNR and binary antipodal (instead of Gaussian) codes was found by Tanaka in [9], and is given by

$$C = \left[\eta \left(\alpha\gamma + \frac{1}{2} \right) - \frac{1}{2} \right] \log_2 e - \frac{1}{2} \log_2 \eta - \alpha \int \log_2 \cosh(z\sqrt{\gamma\eta} + \gamma\eta) Dz \quad \text{bit/s/Hz} \quad (8)$$

where η is the solution to [9]

$$\frac{1}{\eta} = 1 + \alpha\gamma \left(1 - \int \tanh(z\sqrt{\gamma\eta} + \gamma\eta) Dz \right) \quad (9)$$

(we define $Dz \triangleq \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$).

It is not hard to show that, for given γ , the maximum value of C in (8) is also obtained by letting $\alpha \rightarrow \infty$ and coincides with the single-user Gaussian spectral efficiency (7).

Maximum spectral efficiency with optimal coding, separate detection/decoding, and vanishing BER. A common suboptimal practice in multiple-access systems considers separated MUD and single-user decoding. In this case, the decoder is formed by some multiuser detector *front-end*, producing an estimate of the k th user transmit signal for $k = 1, \dots, K$, followed by a bank of K single-user decoders, each processing its own MUD front-end output and producing the decoded message \hat{m}_k independently of the others. The spectral efficiency of such schemes has been examined in several works for various MUD schemes. In [6], Tse and Hanly investigated the multiuser efficiency achieved by *linear* MUD (single-user matched filter (SUMF), linear minimum mean-square error (MMSE), and decorrelating filters) and arbitrary user codes. It is worth noticing that for Gaussian user codes and linear MMSE filter this is the optimal spectral efficiency achievable by separated MUD and decoding, since linear MMSE estimation coincides with the optimal maximum *a posteriori* (MAP) symbol-by-symbol estimation for Gaussian signals. In [17], Müller and Gerstacker found the spectral efficiency with binary user codes and the individually optimal (symbol-by-symbol MAP) MUD front-end. Remarkably, for both Gaussian and binary codes the spectral efficiency under separated MUD and decoding can be written in terms of the corresponding spectral efficiency with joint decoding as [5], [17]

$$C^{\text{Sep}} = C^{\text{Joint}} + \frac{1}{2} \log_2 \eta + \frac{1 - \eta}{2} \log_2 e \quad (10)$$

where C^{Joint} is given by either by (5) or (8) and η is the solution to either (6) or (9), respectively. The term $\frac{1}{2} \log_2 \eta + \frac{1 - \eta}{2} \log_2 e$ quantifies the loss in spectral efficiency due to separation.

Spectral efficiency for given user codes, iterative detection/decoding, and arbitrary target BER. Driven by the success of iterative decoding schemes in single-user channel

¹The achievability results referenced in this section hold under the stronger condition of vanishing message error rate. Well-known converse results ensure that the looser requirement of vanishing BER does not allow any larger rate [3].

coding (see [18] and references therein), “turbo” multiuser joint decoding was proposed in several works (see, for example, [19]–[21] and references in [14]). These algorithms seek a tradeoff between the complexity of optimal joint decoding and the performance loss of separated MUD and single-user decoding. The performance analysis for a wide class of user codes (not necessarily random ensembles) and a class of iterative joint decoders obtained as approximations of the *belief-propagation* (BP) algorithm (see details in Section III) was provided by Boutros and Caire in [14]. This analysis is based on the general technique known as *density evolution* (DE) [22], commonly used to determine the iterative decoding limits of turbo codes and LDPC codes, and is exact in the limit of large block length (notice: to obtain a meaningful large system limit we let first $N \rightarrow \infty$ and then $K \rightarrow \infty$ with $K/L = \alpha$).

Preview of this paper: Several issues are left open in [14]. In particular, how does the BP decoder compare with its interference cancellation (IC)-based approximations? What is the received SNR distribution that maximizes spectral efficiency for given user codes, user target BER and given iterative decoding scheme? How far is the spectral efficiency of an optimized CDMA system with simple (in terms of practical complexity) user codes and iterative joint decoding from the optimal spectral efficiency with optimal (i.e., capacity-achieving) codes and optimal joint decoding? Can we find iterative decoding algorithms with complexity comparable to separated MUD and single-user decoding which still significantly outperforms the separated approach?

In this work, we provide answers to the above questions. In Section II, we recall the BP decoder and some lower complexity approximations based on IC. In Section III, we present the DE analysis of this family of message-passing decoders under random spreading and in the large-system limit. Based on this analysis, in Section IV, we provide a simple linear programming algorithm for the optimization of the received SNR distribution. Our results show that, under constant received SNR, the BP decoder significantly outperforms its IC-based approximations in terms of *power efficiency* (i.e., it requires significantly lower SNR for given target BER). On the other hand, in terms of *spectral efficiency*, the advantage of BP decoding over its approximation based on soft IC and MMSE filtering is only marginal. Moreover, for all the considered decoding algorithms, the spectral efficiency attained under an optimized received SNR distribution is significantly larger than that under constant SNR. Driven by these observations and by the fact that, for the user codes considered here, the optimized received SNR distribution consists of a small number of discrete SNR levels, in Section V we provide a low-complexity approximated version of the MMSE-IC iterative decoder that offers a competitive tradeoff between complexity and performance. Finally, we point out our conclusions in Section VI. The proofs of the main results are provided in the Appendix.

II. ITERATIVE JOINT DECODING ALGORITHM

In the remainder of this work we shall restrict the user codes to be binary antipodal, i.e., $\mathcal{C}_k \subseteq \{-1, +1\}^N$. For a binary

variable c with probability mass function (pmf) ($\Pr(c = +1)$, ($\Pr(c = -1)$) we define its log-ratio by

$$\mathcal{L} \triangleq \log \frac{\Pr(c = +1)}{\Pr(c = -1)}. \quad (11)$$

The BP algorithm [23], [24] approximates iteratively the log-ratios $\mathcal{L}_{k,j}^{\text{bit}}$ corresponding to the marginals of the *a posteriori* joint pmf $\Pr(\mathbf{b}_1, \dots, \mathbf{b}_K | \mathbf{Y})$ of the user information bits. After a given number of iterations, a symbol-by-symbol decision is made according to the threshold rule

$$\hat{b}_{k,j} = \text{sign}(\mathcal{L}_{k,j}^{\text{bit}}). \quad (12)$$

Standard results [24] show that if the *dependency graph* describing $\Pr(\mathbf{b}_1, \dots, \mathbf{b}_K | \mathbf{Y})$ is cycle-free, then BP yields symbol-by-symbol maximum *a posteriori* (MAP) decisions with a finite number of iterations, thus minimizing the BER $P_b^{(k)}$ for each user k . Unfortunately, the dependency graph of the coded multiuser channel (1) has cycles if $K > 1$ and the user codes are nontrivial (i.e., have rate $R_k < 1$). Nevertheless, for sufficiently large block length N and under some randomization of the user codes (e.g., the \mathcal{C}_k 's may be linear convolutional codes the output of which is independently and randomly interleaved before transmission, or LDPC codes whose graph is independently and randomly generated), the probability of finding cycles of any finite-length ℓ decreases linearly with N [14]. Hence, BP decoding is *locally optimal* provided that decisions are made after a finite number ℓ of decoder iterations, while letting N sufficiently large.

The BP iterative joint decoder belongs to the class of *message-passing* decoding algorithms [22]. It is formed by some computation building blocks that exchange *messages* in the form of binary pmfs or, equivalently, of log-ratios. The main building blocks of a BP iterative joint decoder are the soft-input soft-output (SISO) decoders and the individually optimum (IO) MAP multiuser detector (IO-MUD) (see the block diagram in Fig. 1).

SISO decoding is formally expressed by

$$\mathcal{L}_{k,n}^{\text{dec}} = \log \frac{\sum_{\mathbf{c} \in \mathcal{C}_k: c_n = +1} \exp\left(\frac{1}{2} \sum_{j \neq n} c_j \mathcal{L}_{k,j}^{\text{mud}}\right)}{\sum_{\mathbf{c} \in \mathcal{C}_k: c_n = -1} \exp\left(\frac{1}{2} \sum_{j \neq n} c_j \mathcal{L}_{k,j}^{\text{mud}}\right)} \quad (13)$$

for all $k = 1, \dots, K$ and $n = 1, \dots, N$, where $\mathcal{L}_{k,j}^{\text{mud}}$ is the message (log-ratio) sent by the IO-MUD for user k relative to coded symbol $x_{k,j}$ and $\mathcal{L}_{k,n}^{\text{dec}}$ is the so called decoder “extrinsic information” for the coded symbol $x_{k,n}$. For convolutional codes, (13) is efficiently implemented by the well-known forward–backward algorithm [25]. The same forward–backward algorithm can compute the log-ratios $\{\mathcal{L}_{k,j}^{\text{bit}} : j = 1, \dots, B_k\}$ for the user information bits while computing (13).

IO-MUD consists of calculating the *a posteriori* log-ratios with (14) and (15) at the bottom of the following page for all $k = 1, \dots, K$ and $n = 1, \dots, N$. Unfortunately, there is no efficient way to perform this calculation, in general.²

²Based on the fact that $\mathbf{SW}\mathbf{x}$, with $\mathbf{x} \in \{-1, +1\}^K$ is a constellation of N -dimensional points carved from a lattice with generator matrix $\mathbf{M} = \mathbf{SW}$, a modification of the Pohst enumeration of lattice points (sphere decoder [26]) has been proposed by some authors in order to generate a list of candidate transmit vectors and approximate (15) by restricting the sum to a few significant terms in the list [27]. However, this approach is prohibitively complex for large K and/or $\alpha > 1$ (i.e., $K > L$).

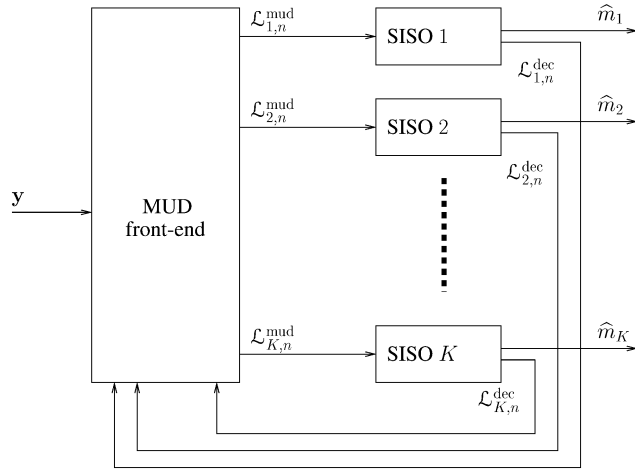


Fig. 1. Block diagram of a multiuser iterative joint decoder.

Various schemes have been proposed to simplify the BP decoder by replacing the IO-MUD block by some simpler soft-in soft-out algorithm. In this work, we shall consider the following options.

Conditional MMSE-IC. The optimal *a posteriori* estimation (15) can be replaced by the unbiased MMSE estimation of $x_{k,n}$ given the received signal \mathbf{y}_n and the SISO decoders extrinsic information $\{\mathcal{L}_{j,n}^{\text{dec}} : j \neq k\}$, given by

$$z_{k,n} = \mathbf{h}_{k,n}^T \left[\mathbf{y}_n - \sum_{j \neq k} w_j \mathbf{s}_j \tanh(\mathcal{L}_{j,n}^{\text{dec}}/2) \right] \quad (16)$$

where the filter $\mathbf{h}_{k,n}$ minimizes the *conditional* MSE as shown by the expression at the bottom of the page under the unbiasedness constraint $w_k \mathbf{h}_{k,n}^T \mathbf{s}_k = 1$ and is given explicitly by

$$\mathbf{h}_{k,n} = \frac{w_k}{\sigma_0^2 \beta_{k,n}} \left[\mathbf{I} + \sum_{j \neq k} \gamma_j (1 - \tanh^2(\mathcal{L}_{j,n}^{\text{dec}}/2)) \mathbf{s}_j \mathbf{s}_j^T \right]^{-1} \mathbf{s}_k \quad (17)$$

where

$$\beta_{k,n} = \gamma_k \mathbf{s}_k^T \left[\mathbf{I} + \sum_{j \neq k} \gamma_j (1 - \tanh^2(\mathcal{L}_{j,n}^{\text{dec}}/2)) \mathbf{s}_j \mathbf{s}_j^T \right]^{-1} \mathbf{s}_k \quad (18)$$

is the output signal-to-interference-plus-noise ratio (SINR).

From (16) and (17), we can write $z_{k,n} = x_{k,n} + \zeta_{k,n}$, where $\zeta_{k,n}$ has mean zero and variance $1/\beta_{k,n}$. Assuming $\zeta_{k,n}$ Gaussian distributed, the log-ratio sent to the SISO decoder is given by

$$\mathcal{L}_{k,n}^{\text{mud}} = 2\beta_{k,n} z_{k,n}. \quad (19)$$

In the large-system limit, the output of the linear MMSE detector converges almost surely to a conditionally Gaussian random variable [8]. Therefore, the Gaussian assumption made in (19) is exact for random spreading CDMA and large K .

The estimator (16) consists of two stages: first, the observation \mathbf{y}_n is rendered zero mean by subtracting the (conditional) mean

$$\begin{aligned} \bar{\mathbf{y}}_{k,n} &= \mathbb{E}[\mathbf{y}_n | \{\mathcal{L}_{j,n}^{\text{dec}} : j \neq k\}] \\ &= \sum_{j \neq k} w_j \mathbf{s}_j \tanh(\mathcal{L}_{j,n}^{\text{dec}}/2). \end{aligned} \quad (20)$$

Then, the linear MMSE estimation of the zero-mean symbol $x_{k,n}$ is obtained by filtering the zero-mean observation $\mathbf{y}_n - \bar{\mathbf{y}}_{k,n}$. This decomposition of linear MMSE estimators is canonical [28]. However, it is interesting to notice that, in this setting, the elimination of the conditional mean of the observation takes on

$$\mathcal{L}_{k,n}^{\text{mud}} = \log \frac{\Pr(x_{k,n} = +1 | \mathbf{y}_n, \mathcal{L}_{1,n}^{\text{dec}}, \dots, \mathcal{L}_{k-1,n}^{\text{dec}}, \mathcal{L}_{k+1,n}^{\text{dec}}, \dots, \mathcal{L}_{K,n}^{\text{dec}})}{\Pr(x_{k,n} = -1 | \mathbf{y}_n, \mathcal{L}_{1,n}^{\text{dec}}, \dots, \mathcal{L}_{k-1,n}^{\text{dec}}, \mathcal{L}_{k+1,n}^{\text{dec}}, \dots, \mathcal{L}_{K,n}^{\text{dec}})} \quad (14)$$

$$= \log \frac{\sum_{\mathbf{c} \in \{\pm 1\}^K : c_k = +1} \exp \left(-\frac{1}{2\sigma_0^2} \left| \mathbf{y}_n - \sum_{j=1}^K w_j \mathbf{s}_j c_j \right|^2 + \frac{1}{2} \sum_{j \neq k} c_j \mathcal{L}_{j,n}^{\text{dec}} \right)}{\sum_{\mathbf{c} \in \{\pm 1\}^K : c_k = -1} \exp \left(-\frac{1}{2\sigma_0^2} \left| \mathbf{y}_n - \sum_{j=1}^K w_j \mathbf{s}_j c_j \right|^2 + \frac{1}{2} \sum_{j \neq k} c_j \mathcal{L}_{j,n}^{\text{dec}} \right)} \quad (15)$$

$$\mathbb{E} \left[\left| x_{k,n} - \mathbf{h}_{k,n}^T \left[\mathbf{y}_n - \sum_{j \neq k} w_j \mathbf{s}_j \tanh(\mathcal{L}_{j,n}^{\text{dec}}/2) \right] \right|^2 \middle| \{\mathcal{L}_{j,n}^{\text{dec}} : j \neq k\} \right]$$

the meaning of *soft IC*. In fact, $\bar{\mathbf{y}}_n$ is the (nonlinear) MMSE estimate of the multiple-access interference $\sum_{j \neq k} w_j x_{j,n} \mathbf{s}_j$ relative to user k , based on the SISO decoder output messages $\{\mathcal{L}_{j,n}^{\text{dec}} : j \neq k\}$.

The above algorithm was introduced in [19], [21] and analyzed in [14]. Since (16) is obtained by solving an MMSE problem conditionally on the SISO decoders extrinsic information and involves soft IC, we shall refer to this detector as the *conditional MMSE-IC* scheme.

Unconditional MMSE-IC. The conditional MMSE-IC detector requires the computation of the filters (17) for each user, each symbol interval, and each decoder iteration. A simplification consists of applying *unconditional*, rather than conditional, linear MMSE estimation to the observation after soft IC. The resulting estimate of $x_{k,n}$ is still given by (16), where the filter $\mathbf{h}_{k,n}$ is replaced by the filter \mathbf{h}_k , minimizing the unconditional mean-squared error (MSE)

$$\mathbb{E} \left[\left\| x_{k,n} - \mathbf{h}_k^T \left[\mathbf{y}_n - \sum_{j \neq k} w_j \mathbf{s}_j \tanh(\mathcal{L}_{j,n}^{\text{dec}}/2) \right] \right\|^2 \right]$$

under the unbiasedness constraint $w_k \mathbf{h}_k^T \mathbf{s}_k = 1$ and is given explicitly by

$$\mathbf{h}_k = \frac{w_k}{\sigma_0^2 \beta_k} \left[\mathbf{I} + \sum_{j \neq k} \gamma_j (1 - \mathbb{E}[\tanh^2(\mathcal{L}_{j,n}^{\text{dec}}/2)]) \mathbf{s}_j \mathbf{s}_j^T \right]^{-1} \mathbf{s}_k \quad (21)$$

where

$$\beta_k = \gamma_k \mathbf{s}_k^T \left[\mathbf{I} + \sum_{j \neq k} \gamma_j (1 - \mathbb{E}[\tanh^2(\mathcal{L}_{j,n}^{\text{dec}}/2)]) \mathbf{s}_j \mathbf{s}_j^T \right]^{-1} \mathbf{s}_k \quad (22)$$

is the output SINR. The log-ratio sent to the SISO decoder is given by (19) with $\beta_{k,n}$ replaced by β_k .

In a practical implementation, the mean $\mathbb{E}[\tanh^2(\mathcal{L}_{j,n}^{\text{dec}}/2)]$ can be replaced by the empirical mean

$$\frac{1}{N} \sum_{n=1}^N \tanh^2(\mathcal{L}_{j,n}^{\text{dec}}/2) \quad (23)$$

that can be computed directly from the output of each j th SISO decoder.

The unconditional MMSE-IC scheme was introduced in [11] and requires the evaluation of only one filter per user per iteration.

Single-user matched filter with IC. A further simplification is obtained by replacing the MMSE filter by the SUMF, and producing an estimate of $x_{k,n}$ as

$$z_{k,n} = \mathbf{s}_k^T [\mathbf{y}_n - \bar{\mathbf{y}}_n].$$

This approach, referred to as the SUMF-IC scheme, was proposed in several early works on uncoded multiuser detection under the name of soft parallel IC (PIC) (see, for example, [29]), and in [20] in the context of iterative joint decoding. It has the advantage of not requiring the computation of matrix inverses. The expression of the output SINR is well known and will be omitted for the sake of brevity.

III. DENSITY EVOLUTION (DE) ANALYSIS

DE consists of propagating through the decoding iterations the probability density of the messages exchanged by a message-passing decoder under the assumption that the messages received at each computation node are statistically independent. Under some mild conditions (notably, that the probability of cycles of any given length ℓ vanishes as the block length N increases), a general concentration theorem [22] ensures that the empirical distribution of the messages at any fixed decoder iteration ℓ converges with probability 1 to the limit density obtained by DE, as $N \rightarrow \infty$. In [14], it is shown that the concentration theorem holds for the coded CDMA channel model and the message-passing decoders presented in the previous section under mild conditions of regularity of the user codes \mathcal{C}_k 's. In particular, the theorem holds for the ensemble of convolutional codes with random independent interleaving, for other standard ensembles of randomlike codes such as LDPC codes and turbo codes.

In the remainder of this paper we make the following assumptions: 1) the user codes are generated randomly and independently in the same ensemble of rate R ; 2) the user spreading sequences \mathbf{s}_k are randomly generated with i.i.d. components according to a symmetric distribution (zero odd moments), variance $1/L$, and finite fourth-order moment; 3) the empirical distribution of the received SNRs, defined by

$$F_\gamma^{(K)}(z) \triangleq \frac{1}{K} \sum_{k=1}^K 1\{\gamma_k \leq z\}$$

converges almost everywhere to a given (nonrandom) distribution $F_\gamma(z)$, as $K \rightarrow \infty$; 4) as anticipated before, we shall study the large-system limit of the iterative decoders by letting first $N \rightarrow \infty$ (to approach the concentration theorem limit) and then $K \rightarrow \infty$ with fixed $K/L = \alpha$ (to remove the randomness due to random spreading). Under these assumptions, the following general result holds.

Proposition 1: For the IO-MUD, conditional MMSE-IC, unconditional MMSE-IC, and SUMF-IC detectors defined above, at each decoder iteration ℓ the log-ratio $\mathcal{L}_{k,n}^{\text{mud}}$ sent to the k th SISO decoder converges in distribution to a Gaussian random variable with conditional mean $2\gamma_k \eta^{(\ell)} x_{k,n}$ (given $x_{k,n} \in \{-1, +1\}$) and variance $4\gamma_k \eta^{(\ell)}$, where the coefficient $\eta^{(\ell)} \in [0, 1]$ depends on the detector and on the iteration ℓ , but it is independent of the user index k . Moreover, $\{\mathcal{L}_{k,n}^{\text{mud}} : n = a, \dots, b\}$ for given finite a and b (that do not depend on the block length N) are asymptotically conditionally independent given the k th user transmitted codeword.

Proof: It follows directly as a corollary of [8], [9], and [13]. \square

Proposition 1 essentially tells that each k th SISO decoder input sequence $\{\mathcal{L}_{k,n}^{\text{mud}} : n = 1, \dots, N\}$, at each decoder iteration ℓ , can be thought as the posterior log-ratio of the output of a *virtual* binary-input additive white Gaussian noise (AWGN) channel $z_{k,n} = x_{k,n} + \zeta_{k,n}$ where $\zeta_{k,n} \sim \mathcal{N}(0, 1/(\gamma_k \eta^{(\ell)}))$. The virtual AWGN channel SNR is $\gamma_k \eta^{(\ell)}$. Hence, $\eta^{(\ell)}$ represents the ratio between the effective SNR for user k at the ℓ th decoder iteration and the nominal received SNR γ_k . Following

the standard definition of [1], we shall refer to $\eta^{(\ell)}$ as the *multiuser efficiency* (ME).

Let us consider the output of the SISO decoder k when its input is driven by the virtual AWGN channel defined above. The probability density function (pdf) of the log-ratio $\mathcal{L}_{k,n}^{\text{dec}}$ defined in (13) satisfies the symmetry condition [30]

$$f(-z) = e^{-z} f(z). \quad (24)$$

In general, $\mathcal{L}_{k,n}^{\text{dec}}$ is non-Gaussian. However, it can be closely approximated by a Gaussian random variable (conditionally on $x_{k,n}$). By imposing the symmetry condition (24) on a Gaussian distribution, we find that the variance must be equal to twice the mean (in absolute value). Therefore, we shall use the *approximation*

$$\mathcal{L}_{k,n}^{\text{dec}} \sim \mathcal{N}\left(2\mu_k^{(\ell)} x_{k,n}, 4\mu_k^{(\ell)}\right). \quad (25)$$

This is equivalent to modeling $\mathcal{L}_{k,n}^{\text{dec}}$ as the posterior log-ratio of the output of a *virtual* binary-input AWGN channel $d_{k,n} = x_{k,n} + \delta_{k,n}$ where $\delta_{k,n} \sim \mathcal{N}\left(0, 1/\mu_k^{(\ell)}\right)$. The above *Gaussian approximation* (GA) has been used extensively to study the performance of turbo codes [31] and LDPC codes [32] under iterative BP decoding.

The output SNR $\mu_k^{(\ell)}$ of the virtual channel defined above depends on the user channel code \mathcal{C} and on the input SNR $\gamma_k \eta^{(\ell)}$. However, since (25) is an approximation, there is some degree of freedom in how to map $\gamma_k \eta^{(\ell)}$ into the corresponding $\mu_k^{(\ell)}$, for a given code \mathcal{C} . We shall use the ‘‘symbol-error rate matching’’ approach proposed in [14]. Namely, let $\epsilon(\text{SNR})$ be the average symbol-error probability at the output of the SISO decoder as a function of the input SNR, defined by

$$\epsilon(\text{SNR}) \triangleq \Pr(\mathcal{L}_{k,n}^{\text{dec}} < 0 | x_{k,n} = +1). \quad (26)$$

Hence, we let

$$\mu_k^{(\ell)} = \left[Q^{-1}\left(\epsilon(\gamma_k \eta^{(\ell)})\right) \right]^2 \quad (27)$$

where $Q(x) \triangleq \int_x^\infty Dz$ is the standard Gaussian tail function.

Suppose that, for a given MUD scheme, we are able to compute $\eta^{(\ell)}$ from the values $\{\mu_k^{(\ell-1)} : k = 1, \dots, K\}$. Then,

the new value $\mu_k^{(\ell)}$ can be computed by (27). The sequence of ME $\{\eta^{(0)}, \eta^{(1)}, \dots, \eta^{(\ell)}, \dots\}$ uniquely defines the evolution of message densities along the decoder iterations (under the GA). Eventually, the DE with GA (referred to as DE-GA in the following) will take on the form of the one-dimensional dynamical system

$$\eta^{(\ell+1)} = \Psi\left(\eta^{(\ell)}\right) \quad (28)$$

where the initial condition $\eta^{(0)}$ and the mapping function Ψ depend on the specific MUD algorithm and on the system parameters, as the channel load α and the limiting distribution of the received SNRs $F_\gamma(z)$. The next propositions give expressions for the mapping function Ψ and for the initial condition $\eta^{(0)}$, for all the MUD algorithms considered.

Proposition 2: The mapping function $\Psi(\eta)$ for the BP decoder is given by the stable solution to the fixed-point equation (29) at the bottom of the page in the interval $[0, 1]$ that minimizes the quantity (30) at the bottom of the page, where $\mathbb{E}_\gamma[\cdot]$ denotes expectation with respect to the received SNR distribution F_γ and where, from (27), we define the function

$$\mu(z) \triangleq \left[Q^{-1}\left(\epsilon(z)\right) \right]^2. \quad (31)$$

Proof: See Appendix A. \square

Equation (29) may have either one (see example in Fig. 2, left) or three distinct solutions (see example in Fig. 2, right) in the interval $[0, 1]$, depending on η , α , and F_γ . If (29) has three solutions $0 \leq \Psi_1 < \Psi_2 < \Psi_3 \leq 1$, Ψ_1 and Ψ_3 are stable fixed points and Ψ_2 is unstable. Then, the desired $\Psi(\eta)$ is given by Ψ_1 or by Ψ_3 for which (30) is minimum.

From the proof given in Appendix A we notice that \mathcal{I} defined in (30) takes on the operational meaning of mutual information per dimension (i.e., spectral efficiency in bits per second per hertz) for the channel (1) where the input symbols \mathbf{x}_n are binary with nonuniform *a priori* marginal pmf given by

$$\Pr(x_{k,n} = +1) = \frac{1 + t_k}{2}$$

$$\frac{1}{\Psi} = 1 + \alpha \mathbb{E}_\gamma \left[\gamma \int_{\mathbb{R}^2} \frac{\left(1 - \tanh^2\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right)\right) \left(1 - \tanh\left(z\sqrt{\gamma\Psi} + \gamma\Psi\right)\right)}{1 - \tanh^2\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right) \tanh^2\left(z\sqrt{\gamma\Psi} + \gamma\Psi\right)} DzDy \right] \quad (29)$$

$$\begin{aligned} \mathcal{I} = & \left[\left(\alpha \mathbb{E}_\gamma[\gamma] + \frac{1}{2} \right) \Psi - \frac{1}{2} \right] \log_2 e - \frac{1}{2} \log_2 \Psi \\ & - \alpha \mathbb{E}_\gamma \left[\int_{\mathbb{R}^2} \left\{ \frac{1 + \tanh\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right)}{2} \log_2 \cosh\left(z\sqrt{\gamma\Psi} + y\sqrt{\mu(\gamma\eta)} + \gamma\Psi + \mu(\gamma\eta)\right) \right. \right. \\ & \left. \left. + \frac{1 - \tanh\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right)}{2} \log_2 \cosh\left(z\sqrt{\gamma\Psi} - y\sqrt{\mu(\gamma\eta)} + \gamma\Psi - \mu(\gamma\eta)\right) \right\} DzDy \right] \\ & - \frac{\alpha}{2} \mathbb{E}_\gamma \left[\int_{\mathbb{R}} \log_2 \left(1 - \tanh^2\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right) \right) Dy \right] \quad (30) \end{aligned}$$

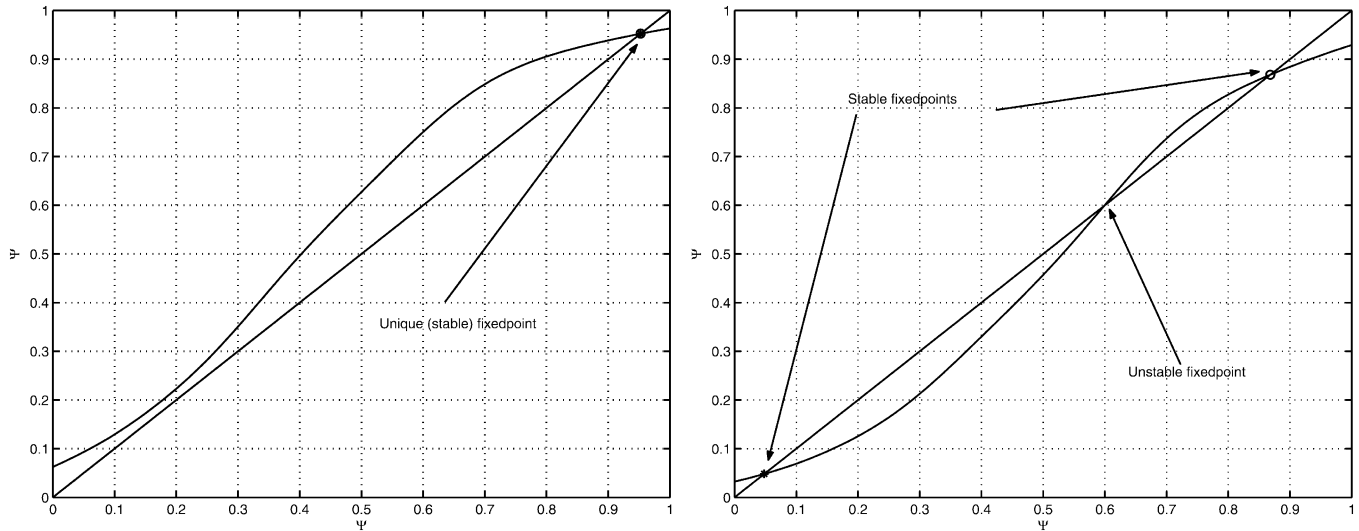


Fig. 2. The curves show the reciprocal of the right-hand side of (29) as a function of $\Psi \in [0, 1]$ for $\alpha = 1.5$ (left), $\alpha = 3.0$ (right) constant $\gamma = 10$ dB and $\eta = 0$. The intersections of the curves with the straight line give the solutions of (29).

(with $t_k \in [-1, 1]$), and where the empirical distribution

$$G_T^{(K)}(z) \triangleq \frac{1}{K} \sum_{k=1}^K 1\{t_k \leq z\}$$

converges almost everywhere as $K \rightarrow \infty$ to the distribution of the random variable $T = \tanh(\mathcal{L}/2)$, with $\mathcal{L} \sim \mathcal{N}(2\mu(\gamma\eta), 4\mu(\gamma\eta))$, and $\gamma \sim F_\gamma$. It is also interesting to notice that the valid solution $\Psi(0)$ of (29) for constant received SNR coincides with the solution of (9), and that, consequently, \mathcal{I} evaluated at $\eta = 0$, $\Psi = \Psi(0)$, and constant received SNR coincides with the spectral efficiency with binary i.i.d. uniform inputs \mathcal{C} given in (8).

For the IC-based iterative decoders we have the following results.

Proposition 3: The mapping function $\Psi(\eta)$ for the conditional MMSE-IC decoder is given by the *unique* solution to (32) (at the bottom of the page) in the interval $[0, 1]$, where $\mu(z)$ is defined in (31).

Proof: See [14, Proposition 3]. \square

Proposition 4: The mapping function $\Psi(\eta)$ for the unconditional MMSE-IC decoder is given by the *unique* solution to (33) (at the bottom of the page) in the interval $[0, 1]$, where $\mu(z)$ is defined in (31).

Proof: The proof of Proposition 4 is given in [11]. For the sake of completeness, we report it in Appendix B. \square

Although not surprising, it is interesting to notice that (32) and (33) reduce to (6) for $\eta = 0$ and constant received SNR. More in general, the solution $\Psi(0)$ of (32) and (33) for $\eta = 0$ and arbitrary F_γ coincide with the ME of linear MMSE MUD found by Tse and Hanly in [6].

Proposition 5: The mapping function $\Psi(\eta)$ for the SUMF-IC decoder is given by

$$\frac{1}{\Psi} = 1 + \alpha \mathbb{E}_\gamma \left[\gamma \left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)) Dy \right) \right] \quad (34)$$

where $\mu(z)$ is defined in (31).

Proof: See Appendix B. \square

For all the above cases, the DE-GA initial condition is given by $\eta^{(0)} = \Psi(0)$, and coincides with the ME of the corresponding MUD scheme used alone, i.e., without coding and iterative decoding. With some effort, it is also possible to verify that, for all $\eta \in [0, 1]$ and all SNR distributions F_γ , the following inequalities hold:

$$0 \leq \Psi_{\text{sumf-ic}}(\eta) \leq \Psi_{\text{unc.mmse-ic}}(\eta) \leq \Psi_{\text{cond.mmse-ic}}(\eta) \leq \Psi_{\text{io-mud}}(\eta) \leq 1. \quad (35)$$

Moreover, for any finite α we have $\eta^{(0)} > 0$ and the functions Ψ are nondecreasing with η . Therefore, the smallest solution of the fixed-point equation $\Psi(\eta) = \eta$ in $[0, 1]$ yields the stable fixed point which the DE-GA tends to, i.e., by letting η^* denote this solution, we have $\lim_{\ell \rightarrow \infty} \eta^{(\ell)} = \eta^*$.

$$\frac{1}{\Psi} = 1 + \alpha \mathbb{E}_\gamma \left[\int_{\mathbb{R}} \frac{\gamma \left(1 - \tanh^2(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)) \right)}{1 + \Psi\gamma \left(1 - \tanh^2(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)) \right)} Dy \right] \quad (32)$$

$$\frac{1}{\Psi} = 1 + \alpha \mathbb{E}_\gamma \left[\frac{\gamma \left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)) Dy \right)}{1 + \Psi\gamma \left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(\gamma\eta)}) + \mu(\gamma\eta) Dy \right)} \right] \quad (33)$$

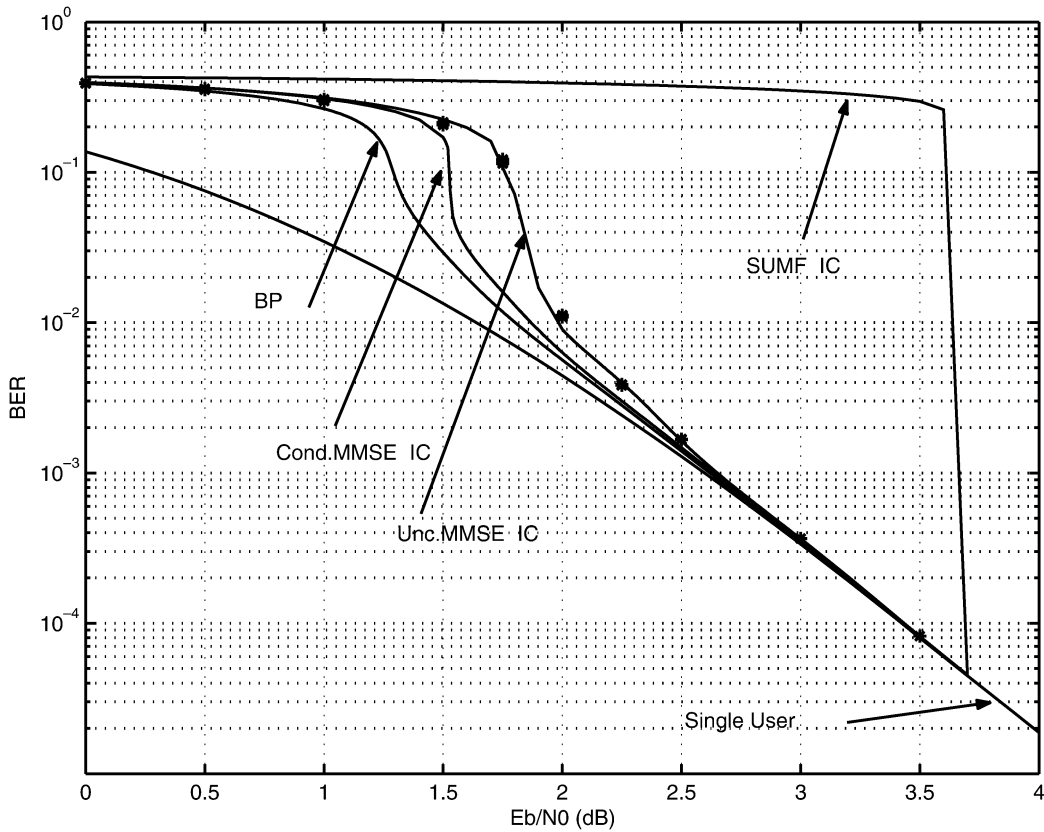


Fig. 3. BER versus E_b/N_0 for $\alpha = 1.0$, constant received SNR, convolutional code with generators $(133, 171)_8$, and different iterative decoding algorithms.

Within the limits of the assumptions made in order to obtain the DE-GA, the iterative decoder performance is completely characterized by the limiting ME η^* . In fact, after many iterations, every k th SISO decoder “sees” a binary-input AWGN channel with SNR $\gamma_k \eta^*$. Therefore, for a given user code \mathcal{C} , the BER is uniquely determined by η^* and by the individual received SNR γ_k . For example, if $\eta^* \approx 1$, every user in the system attains a performance close to its single-user lower bound, as if it were alone in the system. In this case, the iterative decoder is able to *remove* almost entirely the effect of multiple-access interference.

To illustrate the above DE-GA analysis, we computed the BER of a coded CDMA system where the user code \mathcal{C} is the classical 64-state rate-1/2 convolutional code³ with (octal notation [33]) generators $(133, 171)_8$. Figs. 3 and 4 show BER versus E_b/N_0 for constant received SNR, $\alpha = 1.0$ and 2.0, respectively, and various iterative decoding schemes. We notice that the BER shows the typical “waterfall” region (a behavior common to several iterative decoding schemes) where the error curve decreases rapidly with E_b/N_0 and approaches the single-user BER curve. For sufficiently large load α , the waterfall region becomes a “jump,” i.e., an abrupt transition from very large to very small BER. As noticed in [14], this transition corresponds to a *fold bifurcation* [34] of the dynamical system (28) representing the DE-GA. The value of α for which the bi-

furcation appears depends on the decoder algorithm. For example, for $\alpha = 1.0$ (Fig. 3) the SUMF-IC decoder shows the bifurcation behavior while the other detectors have a smooth waterfall. For $\alpha = 2.0$ (Fig. 4) the BP, conditional and unconditional MMSE-IC decoders show bifurcation (at different values of E_b/N_0) while the SUMF-IC decoder is not able to eliminate multiple-access interference (equivalently, the bifurcation appears at infinite E_b/N_0).

IV. RECEIVED SNR DISTRIBUTION OPTIMIZATION

In this section, we aim at optimizing the received SNR distribution F_γ in order to minimize $(E_b/N_0)_{\text{sys}}$ for a given user code \mathcal{C} , desired channel load α , and given maximum BER constraint ϵ (for all users in the system) and for a given iterative decoder in the class of algorithms studied in Sections II and III.

We define the (nonnegative) load density function $\chi(g)$ for $g \in [g_{\min}, g_{\max}]$, with $0 < g_{\min} < g_{\max} < \infty$, such that $\chi(g)dg$ is the fraction of users per chip received at SNR g . In order to stress the dependency of the DE-GA mapping function on $\chi(g)$, we shall use the notation $\Psi(\eta) \equiv \Psi(\chi, \eta)$.

Since the BER is a nondecreasing function of the decoder input SNR, fixing a maximum target BER ϵ to be achieved by all users in the system is equivalent to requiring that the DE-GA fixed point η^* satisfies $\eta^* g_{\min} \geq \text{SNR}(\epsilon)$, where the SNR level $\text{SNR}(\epsilon)$ is determined by the user channel code \mathcal{C} . Suppose that, for some value $0 < \delta < 1$, we can force the DE-GA mapping function such that $\eta^* \geq \delta$. Then, by letting $g_{\min} = \text{SNR}(\epsilon)/\delta$, the target BER is necessarily met by all users in the system. This, however, is just a sufficient condition and might be too

³All the DE-GA numerical results of this paper are based on classical convolutional codes. We evaluated the symbol error probability function $\epsilon(\text{SNR})$ defined in (26) by simulation of the forward-backward decoder in the range of symbol error rate above 10^{-5} and by using the standard union bound on the coded symbol error rate of ML decoding in the region below 10^{-5} .

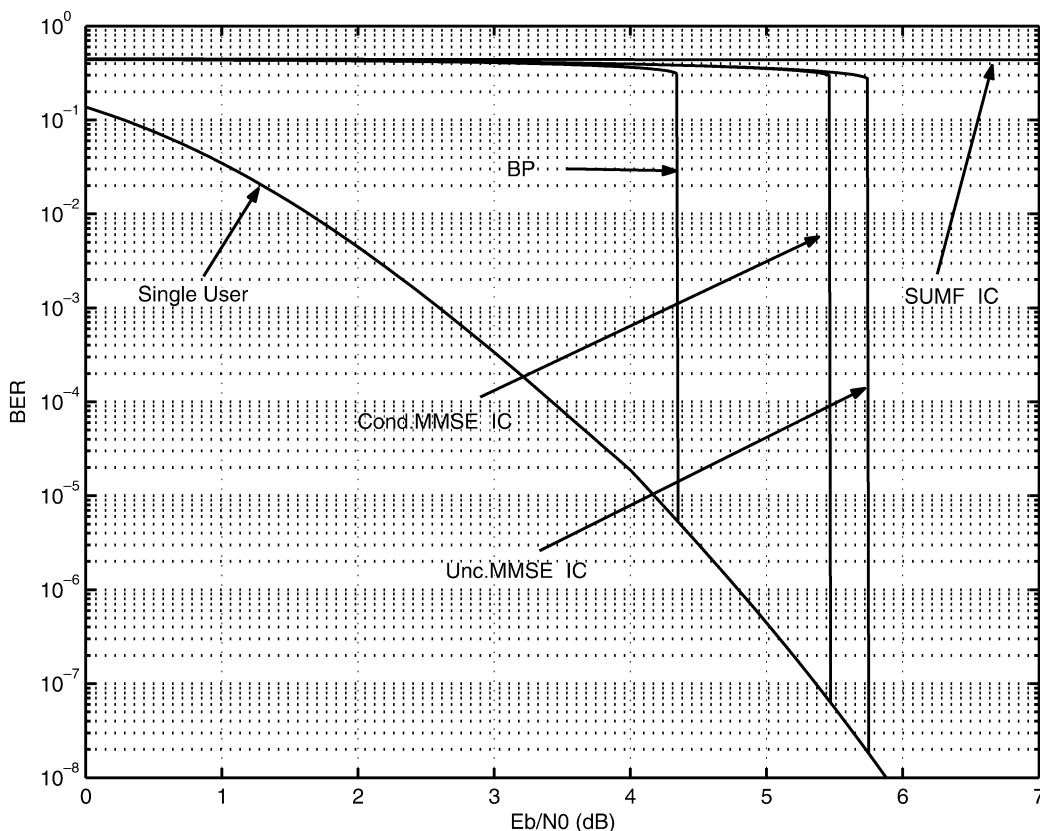


Fig. 4. BER versus E_b/N_0 for $\alpha = 2.0$, constant received SNR, convolutional code with generators $(133, 171)_8$, and different iterative decoding algorithms.

restrictive. Usually, choosing g_{\min} slightly larger than $\text{SNR}(\epsilon)$ yields good results. The value of g_{\max} can be chosen arbitrarily provided that it is sufficiently large. More comments about how to choose g_{\max} are given in the remark on numerical solution of the optimization problem at the end of this section.

We look for the load density $\chi(g)$ solving the optimization program

$$\begin{aligned} & \text{minimize} && \int_{g_{\min}}^{g_{\max}} g\chi(g)dg \\ & \text{subject to} && \begin{cases} \Psi(\chi, \eta) \geq \eta, & \forall \eta \in [0, \delta] \\ \int_{g_{\min}}^{g_{\max}} \chi(g)dg = \alpha \\ \chi(g) \geq 0, & \forall g \in [g_{\min}, g_{\max}]. \end{cases} \end{aligned} \quad (36)$$

Suppose that (36) is feasible and let $\chi^*(g)$ be a solution. Hence, $\chi^*(g)$ has the property of minimizing

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{\int_{g_{\min}}^{g_{\max}} g\chi(g)dg}{2\alpha R}$$

over all load densities $\chi(g)$ such that the spectral efficiency is equal to $\rho = \alpha R$, and the DE-GA has limit $\eta^* \geq \delta$ (implying that all users attain BER not larger than ϵ for what said before). If for some α and ϵ the program (36) is infeasible, then some of the parameters must be changed, for example, by decreasing α , increasing ϵ , or increasing g_{\max} .

Fortunately, for all decoding algorithms considered in this paper, the condition $\Psi(\chi, \eta) \geq \eta$ can be reformulated as a linear constraint with respect to χ . Therefore, (36) is a linear program and can be solved by standard numerical methods (see the remark here below). Before proving the above statement, we would like to point out here that the optimization of $\chi(g)$ via linear programming has striking analogy with the methods for optimizing the degree sequences of LDPC code ensembles, as for example in [32], [35]. However, while a degree distribution is defined over a discrete set of possible degrees, here the load distribution is defined over an interval of possible received SNRs.

For the SUMF-IC decoder, from (34) we rewrite $\Psi(\chi, \eta) \geq \eta$ as (37) (at the bottom of the page) which is clearly linear in $\chi(g)$.

For the conditional and unconditional MMSE-IC decoder, Ψ is given implicitly as the solution of the fixed-point (32) and (33), respectively. These equations have the following property [6]. Let us write (32) and (33) in the form $\Psi = f(\chi, \eta, \Psi)$, and denote by $\Psi(\chi, \eta)$ the solution. Then, for all $x \in [0, 1]$

$$x \leq \Psi(\chi, \eta) \Leftrightarrow x \leq f(\chi, \eta, x). \quad (38)$$

Due to the above iff implication, it follows that $\Psi(\chi, \eta) \geq \eta$ is equivalent to (39) at the top of the following page for conditional MMSE-IC, and to (40), also at the top of the page, for

$$\int_{g_{\min}}^{g_{\max}} g\chi(g) \left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta))Dy\right) dg \leq \frac{1}{\eta} - 1 \quad (37)$$

$$\int_{g_{\min}}^{g_{\max}} g\chi(g) \left\{ \int_{\mathbb{R}} \frac{(1 - \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta)))}{1 + \eta g (1 - \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta)))} Dy \right\} dg \leq \frac{1}{\eta} - 1 \quad (39)$$

$$\int_{g_{\min}}^{g_{\max}} g\chi(g) \frac{(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta)) Dy)}{1 + \eta g (1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta)) Dy)} dg \leq \frac{1}{\eta} - 1 \quad (40)$$

unconditional MMSE-IC. Again, both (39) and (40) are linear constraints in $\chi(g)$.

Finally, for the BP decoder we have to be a bit more careful because of the possibility of multiple solutions to (29) defining the mapping function Ψ . Let us rewrite the fixed-point (29) in the form $\Psi = f(\chi, \eta, \Psi)$, where $f(\chi, \eta, \Psi)$ is the reciprocal of (41) at the bottom of the page, and denote by $\Psi_1(\chi, \eta) \leq \Psi_3(\chi, \eta)$ its stable solutions. Clearly, the inequality $\Psi(\chi, \eta) \geq \Psi_1(\chi, \eta)$ always holds. Thus, the condition $\Psi_1(\chi, \eta) \geq \eta$ for $\eta \in [0, \delta]$ implies the first constraint in (36).

Since $\Psi_1(\chi, \eta)$ is, by definition, the smallest solution of (29), the function $f(\chi, \eta, \psi) - \psi$ is positive for $\psi \in [0, \Psi_1(\chi, \eta))$. Thus, the first constraint in (36) can be replaced by the more stringent constraint

$$f(\chi, \eta, \psi) - \psi > 0, \quad \forall \psi \in [0, \eta), \quad \forall \eta \in [0, \delta]. \quad (42)$$

As desired, (42) is a collection of linear constraints on χ , parameterized by ψ and η .

A simpler approach consists of requiring that (29) has a single solution. Then, if $\Psi(\chi, \eta) = \Psi_1(\chi, \eta)$ is the unique solution of (29), then the condition (38) holds and the corresponding linear constraint is given by (43) at the bottom of the page. By replacing the first constraint in (36) by (43), the solution $\chi^*(g)$ is admissible if $\Psi = f(\chi^*, \eta, \Psi)$ has a unique solution for all $\eta \in [0, 1]$. This can be checked *a posteriori*, i.e., by solving the linear program given by (43), finding a candidate χ^* , and checking the uniqueness of the solution of the fixed-point equation. Fortunately, for practically relevant choices of the code \mathcal{C} and of the target BER ϵ (notably, in all numerical results presented here) we found that the solution of (29) for the candidate optimal χ^* is unique.

Remark: On the numerical solution of the optimization problems. Although the optimization problem (36), formulated as previously given, is linear in the load density function $\chi(g)$, we still have an infinite-dimensional linear program. We can

find an approximated solution by using standard linear programming tools (handling finite-dimensional constraints and objective functions) by 1) discretizing the load density function χ and 2) by imposing the constraint $\Psi(\chi, \eta) \geq \eta$ on a finite and discrete set of values $\{\eta_i\}$.

1) The discretization of the load density function is equivalent to restrict the optimization to densities of the form

$$\chi(g) = \sum_j \alpha_j \delta(g - g_j).$$

Notice that α_j takes on the meaning of the fraction of users per chip received at SNR g_j . For this reason, we shall refer to the α_j 's as the *partial loads*. Clearly, the best results are obtained by using a very fine grid and a large value of g_{\max} , and letting the optimization algorithm find the best load density. Since the objective is minimization of the average SNR, the linear optimization algorithm will associate a zero partial load to values g_j larger than necessary, meaning that no user is required to be received at that high level of SNR. As a matter of fact, numerical experiments show that no further improvement is obtained by increasing g_{\max} beyond a certain value and sampling the g -axis finer than a certain discretization step. The value of g_{\max} and the discretization step of the g -axis depend essentially on the desired spectral efficiency. In particular, the larger α is, the larger the required g_{\max} will be. In the following results, we determined the discretization interval and the value g_{\max} by trial and error.

2) A finite number of linear constraints can be obtained by sampling the interval $[0, \delta]$ in a sufficiently fine grid of points $\{\eta_i\}$, and imposing the constraint $\Psi(\chi, \eta_i) \geq \eta_i$ for all i . In this way, we obtain a finite-dimensional problem where the variables with respect to which optimization is carried out are the partial loads $\{\alpha_j\}$. Some care should be devoted to the fact that a solution satisfying the constraints $\Psi(\chi, \eta_i) \geq \eta_i$ for all i might indeed have some fixed points at some $\eta \in [0, \delta]$ not on the grid

$$1 + \int_{g_{\min}}^{g_{\max}} g\chi(g) \left\{ \int_{\mathbb{R}^2} \frac{(1 - \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta))) (1 - \tanh(z\sqrt{g\Psi} + g\Psi))}{1 - \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta)) \tanh^2(z\sqrt{g\Psi} + g\Psi)} DzDy \right\} dg \quad (41)$$

$$\int_{g_{\min}}^{g_{\max}} g\chi(g) \left\{ \int_{\mathbb{R}^2} \frac{(1 - \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta))) (1 - \tanh(z\sqrt{g\eta} + g\eta))}{1 - \tanh^2(y\sqrt{\mu(g\eta)} + \mu(g\eta)) \tanh^2(z\sqrt{g\eta} + g\eta)} DzDy \right\} dg \leq \frac{1}{\eta} - 1. \quad (43)$$

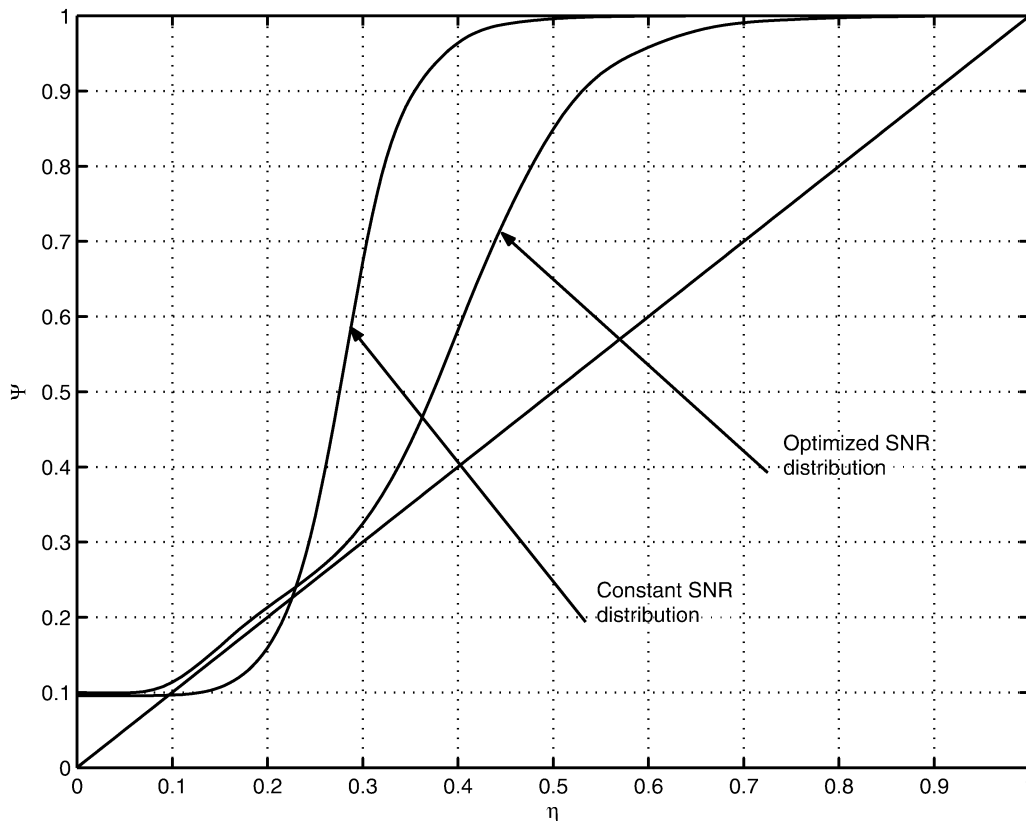


Fig. 5. DE-GA mapping function $\Psi(\mathbf{g}, \alpha, \eta)$ for the BP decoder, $E_b/N_0 = 6$ dB, constant and optimized SNR distributions, $\alpha = 4.5$, and convolutional code with generators $(133, 145, 175)_8$.

$\{\eta_i\}$ or at some η_i when the corresponding constraint is satisfied with equality.

A simple solution to this problem is to enforce the more stringent constraints $\Psi(\chi, \eta_i) \geq \eta_i + \delta'$ for some $\delta' > 0$. The parameter δ' governs the speed of convergence of DE-GA (and eventually of the true iterative decoder) to the fixed point. If δ' is very small, $\Psi(\chi^*, \eta)$ is very close to η for some values of η , and the decoder needs many iterations to find its way out of these “tunnels” (this behavior is completely analogous to what is observed in iterative decoding of turbo codes and LDPC codes through the so-called EXIT diagrams [36]). On the other hand, it may not be possible to obtain small $(E_b/N_0)_{\text{sys}}$ by keeping δ' large. Therefore, δ' can be used as a performance versus complexity tradeoff design parameter. \diamond

As an example of the above optimization technique, consider Fig. 5, showing the DE-GA mapping function $\Psi(\mathbf{g}, \alpha, \eta)$ for the BP decoder, load $\alpha = 4.5$, maximum free distance 64-state rate-1/3 convolutional user codes with generators $(133, 145, 175)_8$ (see [33]), and $(E_b/N_0)_{\text{sys}} = 6$ dB. The curve corresponding to constant received SNR yields $\eta^* \approx 0.1$, i.e., the iterative decoder applied to this system yields very poor performance for all users (10-dB degradation with respect to their single-user performance). On the contrary, the system with optimized SNR distribution yields $\eta^* \approx 1.0$, i.e., each user attains its single-user performance after a sufficiently large number of iterations. The SNR-optimized curve in Fig. 5 is obtained by linear programming by using the constraint (43), enforced over grid of points in $[0, \delta = 0.5]$, equally spaced

TABLE I
OPTIMIZED SNR DISTRIBUTION FOR THE UNCONDITIONAL MMSE-IC RECEIVER WITH $\alpha = 4.5$, CONVOLUTIONAL CODE WITH GENERATORS $(133, 145, 175)_8$, AND $(E_b/N_0)_{\text{sys}} = 6.29$ dB

\mathbf{g} (dB)	α
2.1891	2.5943
5.3141	0.7998
7.1641	1.1063

by 0.01, and by letting $\delta' = 0.01$. For the g -axis we used an equally spaced grid in the decibel scale, between $g_{\min} = 2.18$ dB and $g_{\max} = 20$ dB with a step of 0.005 dB.

Table I gives the optimized SNR distribution achieving $\rho = 1.5$ at $(E_b/N_0)_{\text{sys}} = 6.29$ dB under the unconditional MMSE-IC decoder, at target BER 10^{-5} and convolutional code with generators $(133, 145, 175)_8$. The corresponding channel load is $\alpha = 4.5$ users per chip. The optimized distribution consists of only three discrete SNR levels. Fig. 6 shows the corresponding evolution of the ME provided by the DE-GA. As a further validation of the DE-GA analysis, Fig. 6 shows also the snapshot evolution of the user ME versus decoder iterations for a finite-dimensional system with the same parameters and $L = 32$ ($K = 144$ users). For this simulation, we generated five independent frames of $N = 2000$ coded symbols each, for each frame we generated a set of 144 independently generated spreading sequences with uniform distribution on the 32-dimensional unit sphere. For each iteration, the corresponding

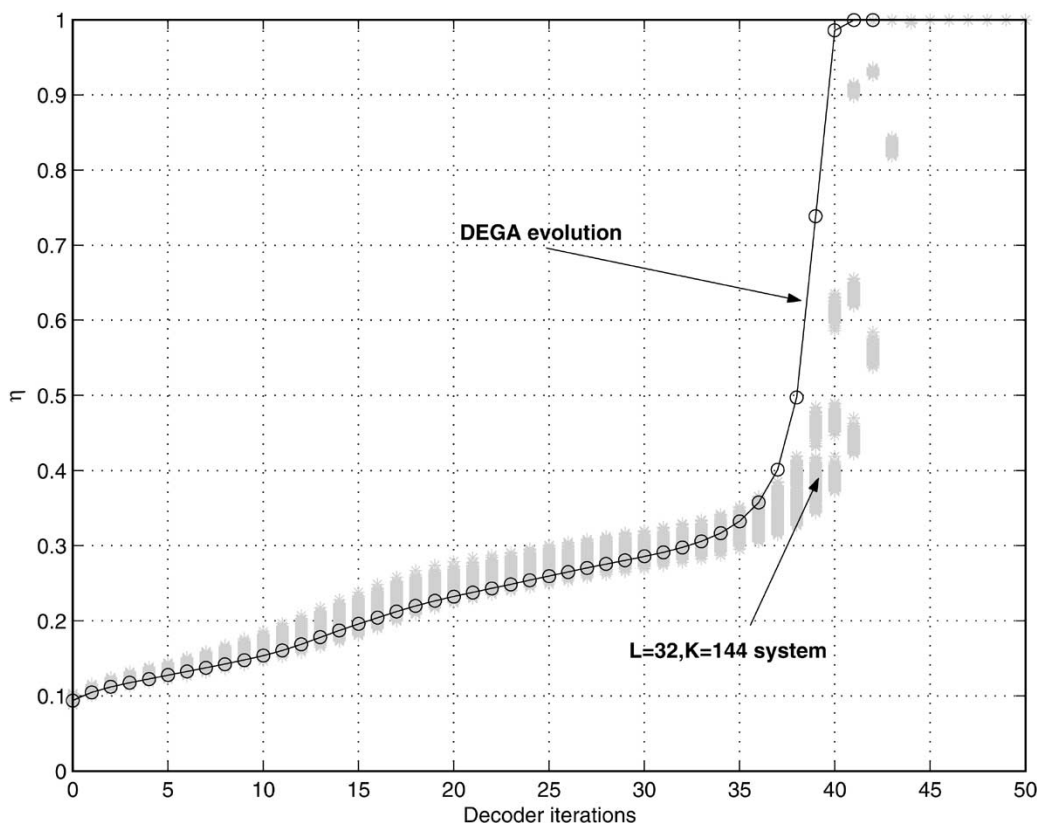


Fig. 6. Evolution of the ME with the decoder iterations for the system of Table I.

cluster of points represents the values of the user ME. We observe that these values are well concentrated around the asymptotic value predicted by the DE-GA analysis. Some discrepancy between the asymptotic analysis and the simulated values is observed only around the point where the ME shows a rapid transition from small values to values close to 1. It is expected that the position of this transition is very sensitive to the residual randomness of the finite-dimensional system. This sensitivity is particularly evident for optimized SNR distributions, characterized by a narrow “tunnel” between the Ψ curve and the main diagonal (see Fig. 5).

Fig. 7 shows the achievable spectral efficiency ρ at target BER 10^{-5} , for coded CDMA systems based on the convolutional code with generators $(133, 145, 175)_8$, and different iterative decoders, with optimized received SNR distribution. For the sake of comparison, we show also the spectral efficiency achievable by optimal Gaussian (or binary) codebooks with joint detection given by (7), with linear MMSE detection (optimal separate detection for Gaussian inputs), and with (suboptimal) linear SUMF detection (these curves have been presented in [4]).

Fig. 8 compares the spectral efficiency ρ at target BER 10^{-5} for the same system described above with the performance of a system with the same user codes and constant received SNR, with iterative detection and with separate detection [51] (corresponding to the performance of the iterative decoders after the first iteration).

Based on these results, the following remarks are in order.

- All spectral efficiencies of the convolutionally coded systems are zero for $(E_b/N_0)_{\text{sys}} < 3.94$ dB, that is, the value of E_b/N_0 needed for a single user to achieve BER $= 10^{-5}$. This limit depends on the user code alone, and can be improved by choosing a more powerful code.
- As said in Section I, for both Gaussian and binary inputs, spectral efficiency is maximized by infinite load and vanishing per-user rate. On the contrary, the spectral efficiency curves for the convolutionally coded CDMA system with iterative multiuser joint decoding reported in Figs. 7 and 8 correspond to per-user rate $R = 1/3$ bit/symbol and finite α users per chip. In this sense, these curves are much more meaningful from the viewpoint of practical CDMA design.
- For large $(E_b/N_0)_{\text{sys}}$, the iteratively decoded systems with optimized SNR distribution are not interference limited, in the sense that their spectral efficiency increases with $(E_b/N_0)_{\text{sys}}$. Remarkably, for the BP and the MMSE-IC decoders, the large- $(E_b/N_0)_{\text{sys}}$ slope of spectral efficiency is (close to) optimal, at least in the considered range of $(E_b/N_0)_{\text{sys}}$.
- CDMA systems with constant received SNR are basically interference limited, and iterative joint decoding provides a significant gain with respect to conventional separate multiuser detection and single-user decoding only for small $(E_b/N_0)_{\text{sys}}$.
- The unconditional MMSE-IC yields spectral efficiency very close to BP with much smaller complexity with respect to both BP and conditional MMSE-IC. This makes the unconditional MMSE-IC decoder a good candidate for

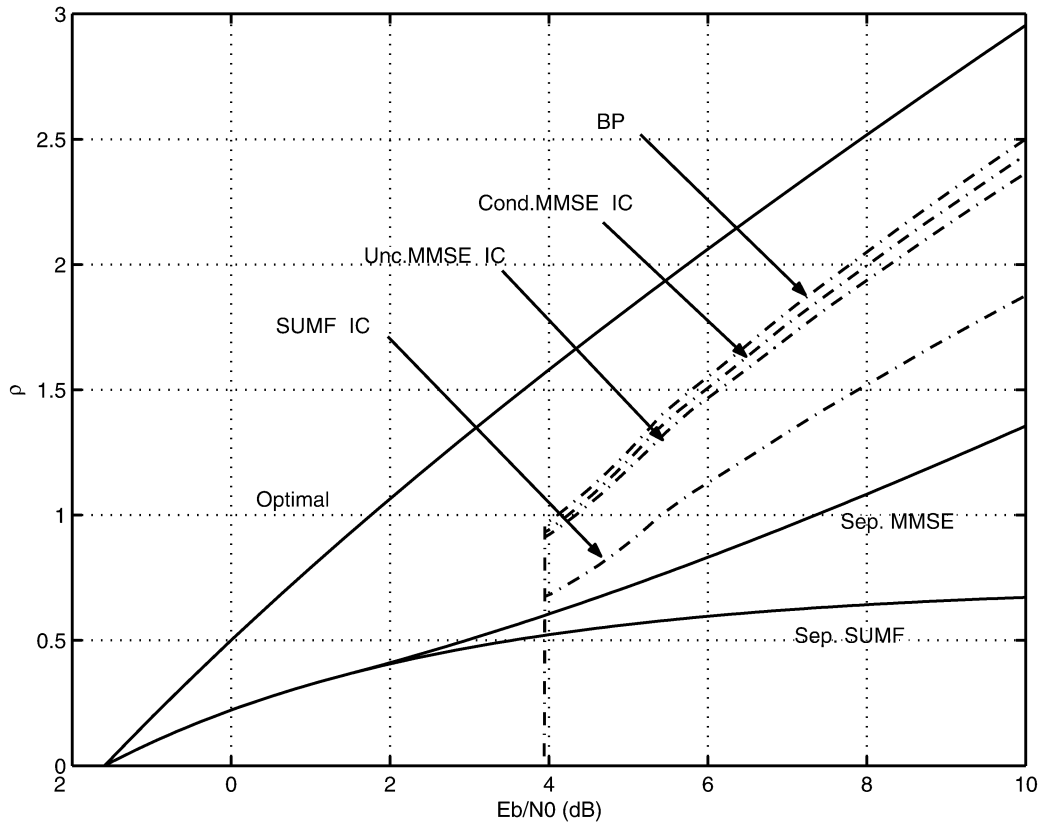


Fig. 7. Spectral efficiency versus $(E_b/N_0)_{sys}$ at $BER \leq 10^{-5}$ for convolutionally coded CDMA with user codes with generators $(133, 145, 175)_8$, optimized received SNR distribution, and different iterative decoding algorithms. Curves for joint detection and optimal codes (binary and Gaussian input), and separated MMSE and SUMF detection and decoding (Gaussian inputs) are reported for the sake of comparison.

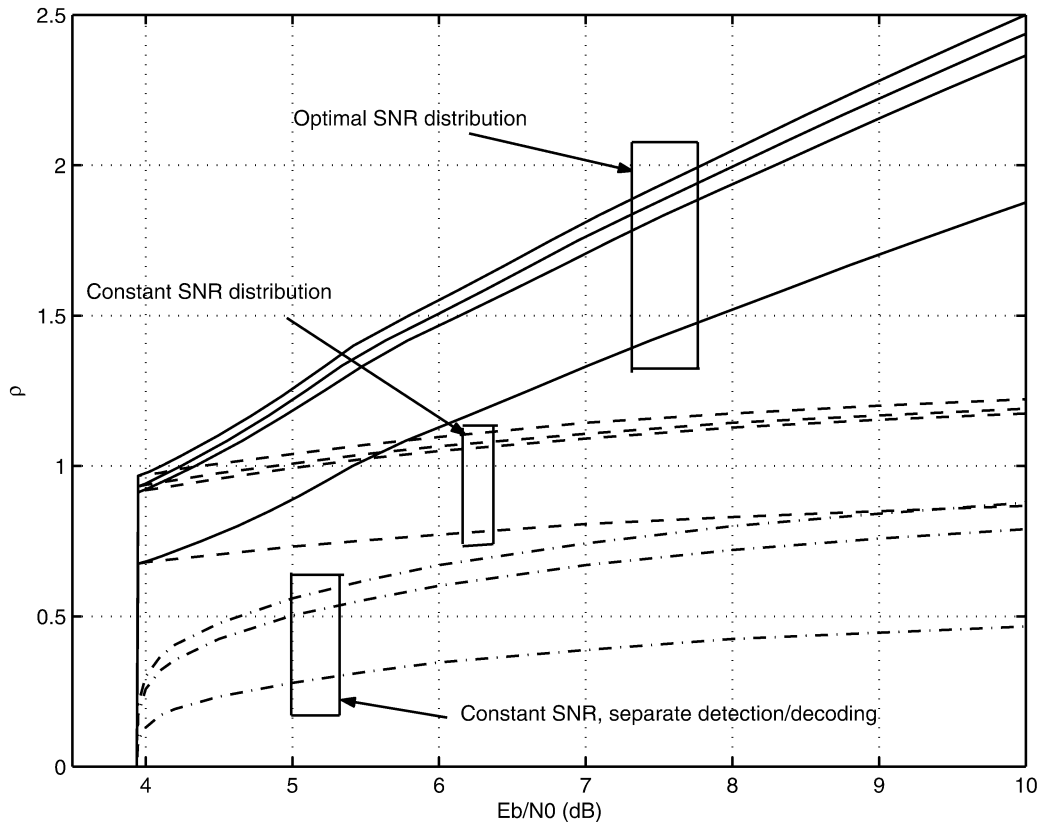


Fig. 8. Spectral efficiency versus $(E_b/N_0)_{sys}$ at $BER \leq 10^{-5}$ for convolutionally coded CDMA with user codes with generators $(133, 145, 175)_8$, optimized received SNR distribution, and constant received SNR distribution with iterative and separate detection and decoding. Each set of curves shows the performance of BP, conditional MMSE-IC, unconditional MMSE-IC, and SUMF-IC (curves from top to bottom). Obviously, for separate detection conditional and unconditional MMSE coincide.

high-performance low-complexity iterative multiuser decoding. This point will be elaborated further in Section V.

V. LOW-COMPLEXITY IMPLEMENTATION

In this section, we consider the computational complexity of the iterative multiuser joint decoders analyzed previously. Our considerations apply to the so-called periodic random spreading, i.e., where the user spreading sequences are randomly generated and used for a long sequence of codewords (blocks of N symbols).⁴

The complexity per information bit is given essentially by three operations: the calculation of the multiuser detectors (e.g., the conditional or unconditional MMSE filters), the activation of the MUD front-end and the activation of the SISO decoders (see Fig. 1). We assume that the computation load of the SISO decoders is linear⁵ with N , therefore it contributes with an additive constant to the overall complexity per information bit, independently of the MUD scheme used. The complexity per information bit of MUD front-end calculation and activation depends on the number of users K , the number of iterations I , and the block length N . The spreading factor L is constrained by the number of users, since $K = \alpha L$, and it is not a free parameter for fixed load.

The IO-MUD and the SUMF-IC decoders do not require any front-end calculation. While the SUMF-IC front end has activation complexity $O(IK)$ (corresponding to one matrix–vector product per symbol per iteration, with complexity $O(IK^2N)$, divided by the total number of information bits per block, equal to KNR), the IO-MUD front end has activation complexity $O(I^2K)$. The activation complexity of the conditional and unconditional MMSE-IC front ends is equivalent to that of the SUMF, namely, $O(IK)$. In fact, any linear front end corresponds to a matrix–vector product per symbol per iteration. However, these MMSE-IC variants have very different front-end calculation complexity: conditional MMSE-IC requires the computation of a matrix inverse per symbol per iteration (with complexity per information bits $O(IK^2)$), while the unconditional MMSE-IC requires the computation of a matrix inverse per iteration (with complexity per information bits $O(IK^2/N)$). This clearly shows the advantage of unconditional MMSE-IC.

Nevertheless, while it is clear that for $N \gg K$ the evaluation complexity of the unconditional MMSE-IC front end becomes negligible; in practice, it might still be too large with respect to the SUMF.⁶ Motivated by this fact, in the following we propose a suboptimal scheme that closely approximates the performance

of the unconditional MMSE-IC while yielding a front-end calculation complexity that is constant with the number of iterations.

One crude method to obtain front-end calculation complexity that is constant with respect to I , reported in the literature [41], consists of using the standard linear MMSE detector for the first few iterations, and, assuming that the decoder is able to eliminate multiple-access interference, switch to the standard SUMF filter when the residual interference symbol variances are below a certain threshold. This approach does not take into account the fact that, under the optimal SNR distribution, users are received at different SNR levels and the evolution of their residual symbol variances during the decoder iterations may be very different. Indeed, we may expect that users received at higher SNR levels are correctly estimated and canceled much faster than users received at low SNR. Thus, the stronger the interferers are at the beginning, the weaker they are after some iterations. This is particularly problematic for linear detectors in overloaded systems, i.e., $\alpha > 1$, where linear detectors focus their interference suppression capabilities on those subspaces with strong interferers while neglecting subspaces with moderate to weak interfering users. If the choice of the subspace is not updated during iterations, the linear multiuser detector becomes severely mismatched and even worse in performance than no multiuser detector at all.

Empirical evidence was found in the previous section that optimal received power distributions for many practical codes and a wide range of system parameters such as load and $(E_b/N_0)_{\text{sys}}$ consist of only a small number J of different power levels (see Table I for an example resulting in $J = 3$). Since there are only J different classes of users, there is reason to assume that switching the subspace which the linear multiuser detector is focusing onto for only J times is sufficient and not far in performance from switching I times, i.e., after every iteration. If $J \ll I$, a significant reduction in complexity is possible. Often, the number of iterations I is larger than the number of power classes by one or more orders to magnitude, see Fig. 6 for an example.

In order to exploit the fact that there are only J different groups of users to distinguish, we propose to switch the subspace only J times during all iterations, but to update the weighting of the interference suppression capabilities versus the noise suppression capabilities—a process which can be done at insignificant complexity—on an iteration-by-iteration basis. This means that at the beginning of the iterations, we process all users by an unconditional MMSE filter which does not adapt to the changes of the power distribution of the interference, but only to the changes of the total interference power. If, after some iterations, the total interference of one class of users has become insignificant, the MMSE filter is recalculated neglecting the presence of interference from these users. Thus, the second MMSE filter to be calculated mitigates only interference from the remaining $J - 1$ classes of users. In order to enable an easy adaptation to the diminishing total interference power, we make use of the singular value decomposition. The switch points, i.e., the points at which iteration a new MMSE filter should be calculated, are determined by approximating the performance of the mismatched MMSE

⁴We hasten to say that rather different approaches based on matrix polynomials should be considered for low-complexity algorithms in the case of aperiodic random spreading, where a new set of spreading sequences is used on every symbol interval (see [10], [37]–[40]).

⁵This assumption is verified for convolutional, LDPC, or turbo user codes.

⁶We have also to take into account that, for some implementation architectures, while the linear front-end activation is implemented by very simple dedicated hardware such as a bank of correlators, the calculation of the MMSE filters requires a matrix inverse and must be performed at higher precision. Therefore, the effective impact of the front-end calculation on the overall complexity might be much larger than the simple operation count may reveal.

filter by means of large system analysis. This requires to solve a one-dimensional fixed-point equation per iteration, a task with negligible complexity even compared to the SUMF.

In the following, we will first focus on the calculation of the filter assuming we know when we should switch from one to another and then, given the structure of the filters, discuss appropriate switching criteria. For the purpose of finding the switch criterion, it will be a prerequisite to study the large-system performance of mismatched linear MMSE detectors.

A. Filter Calculation

Consider again the CDMA channel model (1) and an IC-based iterative decoder that, at a certain iteration ℓ , produces the n th observable for the SISO decoder of user k as

$$z_{k,n}^{(\ell)} = \left(\mathbf{f}_k^{(\ell)} \right)^T \left(\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{t}_n^{(\ell)} + \mathbf{s}_k w_k t_{k,n}^{(\ell)} \right) \quad (44)$$

where $\mathbf{t}_n^{(\ell)} = \left(t_{i,n}^{(\ell)}, \dots, t_{K,n}^{(\ell)} \right)^T$ and where we define

$$t_{j,n}^{(\ell)} = \tanh \left(\mathcal{L}_{j,n}^{\text{dec},\ell} / 2 \right)$$

to be the current estimate of the j th user symbol given the SISO decoders output messages at the current iteration (see (20)), and where $\mathbf{f}_k^{(\ell)}$ is an appropriately chosen filter.⁷

The unbiased unconditional MMSE criterion leads to (21), that it is rewritten by using simple matrix identities as

$$\mathbf{f}_k^{(\ell)} = \frac{\left[\mathbf{S}\mathbf{T}\mathbf{V}^{(\ell)}\mathbf{S}^T + \mathbf{I} \right]^{-1} \mathbf{s}_k}{\mathbf{s}_k^T \left[\mathbf{S}\mathbf{T}\mathbf{V}^{(\ell)}\mathbf{S}^T + \mathbf{I} \right]^{-1} \mathbf{s}_k} \quad (45)$$

where we define $\mathbf{\Gamma} \triangleq \text{diag}(\gamma_1, \dots, \gamma_K)$ and the residual symbol covariance matrix at iteration ℓ as

$$\begin{aligned} \mathbf{V}^{(\ell)} &\triangleq \mathbb{E} \left[\left(\mathbf{x}_n - \mathbf{t}_n^{(\ell)} \right) \left(\mathbf{x}_n - \mathbf{t}_n^{(\ell)} \right)^T \right] \\ &= \text{diag} \left(v_1^{(\ell)}, \dots, v_K^{(\ell)} \right) \end{aligned} \quad (46)$$

with

$$v_k^{(\ell)} = 1 - \mathbb{E} \left[\left(t_j^{(\ell)} \right)^2 \right], \quad k = 1, \dots, K. \quad (47)$$

Notice that $\mathbf{V}^{(\ell)}$ is *exactly* diagonal in the limit for large block length N because $t_{j,n}^{(\ell)}$ are obtained from the SISO decoders *extrinsic information* [14].

Under the optimized received SNR distribution, we shall assume that the users are grouped into $J \ll K$ classes of sizes K_1, \dots, K_J . User k in class j is received at SNR level $\gamma_k = g_j$. As in Section IV, we let $g_1 \leq \dots \leq g_J$ and enumerate the users such that users $k \in \{ \mathcal{K}_{j-1} + 1, \dots, \mathcal{K}_j \}$ belong to class j , where $\mathcal{K}_0 \triangleq 0$ and $\mathcal{K}_j \triangleq \sum_{i=1}^j K_i$.

The proposed low-complexity approach makes subsequently use of J different linear detectors (by different detector, we mean a newly calculated MMSE filter, not just an adaptation

⁷We use $\mathbf{f}_k^{(\ell)}$ instead of \mathbf{h}_k as in Section II in order to stress the fact that here the filter does not necessarily coincide with the unconditional MMSE filter (21). Moreover, we specify explicitly the iteration index ℓ since it is relevant in the definition of the low-complexity algorithm.

to the decreasing total interference power). Detector number j at iteration ℓ assumes user SNRs given by

$$u_{j,k}^{(\ell)} = \begin{cases} \xi^{(\ell)} \gamma_k v_k^{(\ell_j)}, & \text{for } k = 1, 2, \dots, \mathcal{K}_j \\ 0, & \text{for } k = \mathcal{K}_j + 1, \mathcal{K}_j + 2, \dots, K \end{cases} \quad (48)$$

where $\xi^{(\ell)}$ is an iteration-dependent scaling factor common to all users that accounts for the diminishing total interference power and ℓ_j is an iteration index that characterizes the j th detector. In matrix form, we define the diagonal matrix \mathbf{Z}_j such that its k th diagonal element is zero if k belongs to a class larger than j and one otherwise, and let the diagonal matrix of *nominal* received SNRs for the j th detector be given by

$$\mathbf{U}_j^{(\ell)} = \xi^{(\ell)} \mathbf{\Gamma} \mathbf{V}^{(\ell_j)} \mathbf{Z}_j. \quad (49)$$

Equation (49) is meaningful only for $\ell \geq \ell_j$. As it will be clear in the following, detectors are used in the order $j = J, J-1, \dots, 1$, and the indices ℓ_j determine the detector switch points, i.e., the j th detector is used for $\ell = \ell_j, \dots, \ell_{j-1} - 1$, where $\ell_J = 0$ and ℓ_0 is the maximum number of iterations.

In order to obtain a computationally efficient form for the j th detector, we decompose the spreading matrix as

$$\mathbf{S} = \mathbf{S}\mathbf{Z}_j + \underbrace{\mathbf{S}(\mathbf{I} - \mathbf{Z}_j)}_{\triangleq \tilde{\mathbf{S}}_j}. \quad (50)$$

Note that $\mathbf{S}\mathbf{Z}_j$ contains the sequences of all users to be processed by the multiuser detector, while $\tilde{\mathbf{S}}_j$ contains the sequences of all users to be processed by a SUMF only. We replace the *true* SNR diagonal matrix $\mathbf{\Gamma} \mathbf{V}^{(\ell)}$ in (45) by $\mathbf{U}_j^{(\ell)}$. We introduce the singular value decomposition

$$\mathbf{S} \left(\mathbf{\Gamma} \mathbf{V}^{(\ell_j)} \mathbf{Z}_j \right)^{1/2} = \mathbf{\Phi}_j \mathbf{D}_j \mathbf{\Theta}_j^T \quad (51)$$

such that $\mathbf{\Phi}_j$ and $\mathbf{\Theta}_j$ are unitary and \mathbf{D}_j is diagonal up to some additional columns or rows which are all zero. We define

$$\begin{aligned} \mathbf{Q}_j &\triangleq \mathbf{\Phi}_j^T \mathbf{S} \\ &= \mathbf{D}_j \mathbf{\Theta}_j^T \mathbf{\Gamma}^{-1/2} \left(\mathbf{V}^{(\ell_j)} \right)^{-1/2} + \mathbf{\Phi}_j^T \tilde{\mathbf{S}}_j. \end{aligned} \quad (52)$$

Note that, though (53) looks more complicated, it may require fewer computing effort than (52) due to the diagonal structure of the matrices $\mathbf{D}_j, \mathbf{\Gamma}$, and $\mathbf{V}^{(\ell_j)}$ and the zero columns in $\tilde{\mathbf{S}}_j$. By using (52) and (49) in (45) and (44), we can write the j th detector filter for user k at iteration $\ell = \ell_j, \dots, \ell_{j-1} - 1$ as

$$\tilde{\mathbf{f}}_{j,k}^{(\ell)} = \frac{\mathbf{\Phi}_j \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1} \mathbf{q}_{j,k}}{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1} \mathbf{q}_{j,k}} \quad (54)$$

where $\mathbf{q}_{j,k}$ denotes the k th column of \mathbf{Q}_j , and the n th output of the detector as

$$\begin{aligned} \tilde{z}_{k,n}^{(\ell)} &= \frac{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1} \mathbf{\Phi}_j^T}{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1} \mathbf{q}_{j,k}} \\ &\quad \times \left(\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{t}_n^{(\ell)} + \mathbf{s}_k w_k t_{k,n}^{(\ell)} \right) \\ &= \frac{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1}}{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1} \mathbf{q}_{j,k}} \\ &\quad \times \left(\mathbf{\Phi}_j^T \mathbf{y}_n - \mathbf{Q}_j \mathbf{W}\mathbf{t}_n^{(\ell)} + \mathbf{q}_{j,k} w_k t_{k,n}^{(\ell)} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1}}{\mathbf{q}_{j,k}^T \left[\xi^{(\ell)} \mathbf{D}_j^2 + \mathbf{I} \right]^{-1} \mathbf{q}_{j,k}} \\
&\quad \times \underbrace{\left(\Phi_j^T \mathbf{y}_n - \mathbf{Q}_j \mathbf{W} \mathbf{t}_n^{(\ell)} \right)}_{\triangleq \mathbf{d}_{n,j}^{(\ell)}} + w_k t_{k,n}^{(\ell)}. \quad (55)
\end{aligned}$$

The front-end calculation of the proposed scheme requires a singular value decomposition (51) and a matrix multiplication (52) or (53) at each switching point. Thus, the overall calculation complexity of the proposed approach is $O(JK^2/N)$. As anticipated before, our approach is very effective for small J and a large number of decoder iterations (typical of heavily loaded systems attaining large spectral efficiency).

B. Mismatched MMSE Detection

Two questions have been left open: how to determine the detector switch points ℓ_j and how to choose the scaling factor $\xi^{(\ell)}$. Before we can solve these issues, we need some results on the performance of mismatched linear MMSE detectors.

For the time being, let $\xi^{(\ell)}$ be a given function of the iteration index and of the other system parameters (including the received signal block \mathbf{Y}) that can be easily computed in real time along the decoder iterations. The filter $\tilde{\mathbf{f}}_{j,k}^{(\ell)}$ can be regarded as a *mismatched* MMSE filter that assumes user received SNRs given by the diagonal elements of $\mathbf{U}_j^{(\ell)}$ rather than the exact values given by $\mathbf{\Gamma V}^{(\ell)}$. In order to determine an effective switching criterion, we make use of the following result characterizing the ME of an MMSE filter with power mismatch:

Proposition 6: Consider the CDMA system with K users and spreading factor L defined by $\mathbf{y} = \mathbf{S P}^{1/2} \mathbf{x} + \mathbf{v}$ where $\mathbf{v} \sim \mathcal{N}(0, \mathbf{I})$ and the usual assumptions on random spreading made in Section III hold. Let $\mathbf{P} \triangleq \text{diag}(P_1, \dots, P_k)$ such that $\max_k P_k \leq \bar{P}$, and let (U_1, \dots, U_k) be an arbitrary sequence of positive numbers such that $\min_k U_k \geq \underline{U}$, where \bar{P} and \underline{U} are fixed, finite, and positive constants independent of K . Assume that, as $K \rightarrow \infty$, the joint empirical distribution of the pairs (P_k, U_k) , defined by

$$F^{(K)}(p, u) \triangleq \frac{1}{K} \sum_{k=1}^k \mathbf{1}\{P_k \leq p, U_k \leq u\}$$

converges almost everywhere to a given distribution $F_{P,U}(p, u)$. Then, by letting $K \rightarrow \infty$ with $K/L = \alpha$, the ME of a linear detector obtained as the MMSE filter assuming received user powers given by $\{U_k\}$ instead of the true values $\{P_k\}$ converges almost surely for all user K to the value κ given by

$$\kappa = \eta \frac{1 + \alpha \mathbb{E}_U \left[\frac{U}{(1 + \eta U)^2} \right]}{1 + \alpha \mathbb{E}_{P,U} \left[\frac{P}{(1 + \eta U)^2} \right]} \quad (56)$$

where η is the solution to

$$\frac{1}{\eta} = 1 + \alpha \mathbb{E}_U \left[\frac{U}{1 + \eta U} \right] \quad (57)$$

and where $\mathbb{E}_U[\cdot]$ and $\mathbb{E}_{P,U}[\cdot]$ denote expectations with respect to $(P, U) \sim F_{P,U}(p, u)$.

Proof: The proof of the above proposition follows as a corollary of the proof of the main result of [8]. A concise and self-contained sketch of proof (skipping technicalities on almost-sure convergence) is given in Appendix C. \square

Note that η given in (57) is the ME of an MMSE filter matched to powers $\{U_k\}$ rather than $\{P_k\}$. Hence, it can be regarded as the *nominal* ME of the mismatched detector, while κ given by (56) is the *true* ME of the mismatched detector. We say that the nominal powers $\{U_k\}$ are *adapted* to the actual powers $\{P_k\}$ if the two sequences $\{P_k\}$ and $\{U_k\}$ can be sorted in nondecreasing order by the same permutation and have the same sum. We say that $\{U_k\}$ is a *conservative* choice of the nominal powers if it is adapted to $\{P_k\}$ and yields $\kappa \geq \eta$, i.e., the actual detector performance is better than what can be expected if the users had powers $\{U_k\}$ instead of $\{P_k\}$. The following result gives a sufficient condition for a conservative nominal power assignment.

Corollary 6.1: Let $\{U_k\}$ be adapted to $\{P_k\}$. Then, in the limit for large K , $\{U_k\}$ is a conservative choice of the nominal powers if $\{P_k\}$ majorizes $\{U_k\}$.⁸

Proof: See Appendix C. \square

Corollary 6.1 generalizes the result of [10] that, rephrased in our terminology, states that

$$\left\{ U_k = \frac{1}{K} \sum_{j=1}^K P_j : k = 1, \dots, K \right\}$$

is a conservative nominal power assignment for any received power sequence $\{P_k\}$, since it is well known that any $\{P_k\}$ majorizes its corresponding constant mean value sequence [42].

C. Switch Point Calculation

A sensible criterion to switch from detector j to detector $j-1$ at a given iteration ℓ consists of choosing the detector with largest ME. Let j be the current detector index. At each iteration ℓ , we decide if using detector j or switching to detector $j-1$ by comparing their MEs $\kappa_j^{(\ell)}$ and $\kappa_{j-1}^{(\ell)}$ via the large-system formula (56). For detector j we have, from Proposition 6, (58) at the bottom of the following page, while for detector $j-1$ we have (59) at the bottom of the following page. If $\kappa_j^{(\ell)} < \kappa_{j-1}^{(\ell)}$, then ℓ_{j-1} is set equal to ℓ and the detector is switched from j to $j-1$.

In writing (59) we implicitly assumed that $\xi^{(\ell)} = 1$ at each switch point (in particular, for $\ell = \ell_{j-1}$). A very effective heuristic choice enforcing the condition $\xi^{(\ell_j)} = 1$ for all $j = J, J-1, \dots, 1$ is given by

$$\xi^{(\ell)} = \frac{\sum_{k=1}^K v_k^{(\ell)}}{\sum_{k=1}^K v_k^{(\ell_j)}}, \quad \text{for } \ell = \ell_j, \ell_j + 1, \dots, \ell_{j-1} - 1. \quad (60)$$

Optimizing the detector with respect to $\xi^{(\ell)}$ appears to be a hard problem and remains open.

⁸A sequence $\{P_k\}$ is said to majorize [42] a sequence $\{U_k\}$ if

$$\sum_{j=1}^k U_{(j)} \geq \sum_{j=1}^k P_{(j)}, \quad \forall K = 1, \dots, k$$

with equality for $k = K$, where $\{P_{(k)}\}$ and $\{U_{(k)}\}$ denote the nondecreasing arrangements of $\{P_k\}$ and $\{U_k\}$, respectively.

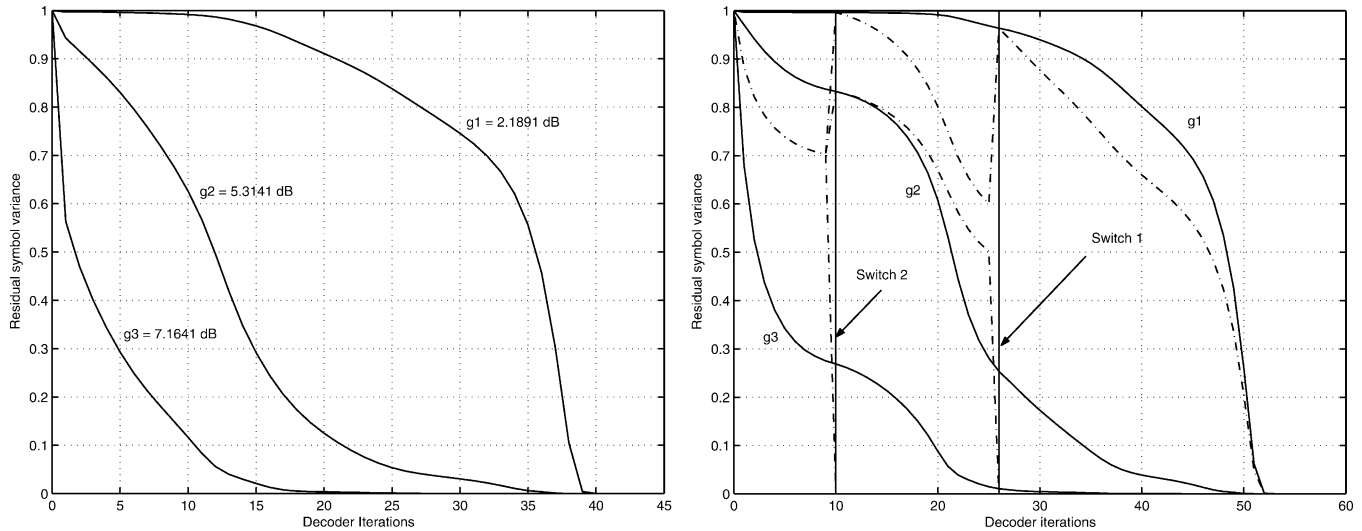


Fig. 9. Evolution of the user residual symbol variances with the decoder iterations for the system of Table I with unconditional MMSE-IC iterative decoder (left) and the low-complexity detector algorithm (right).

D. Performance

In the following, we compare the performance of the proposed algorithm to the performance of unconditional MMSE-IC. Consider again the system and SNR distribution of Table I. The distribution is composed by only $J = 3$ SNR levels, denoted by g_1, g_2, g_3 . Fig. 9 (left) shows the evolution of the residual user symbol variances for the three classes of users versus the decoder iterations. We notice that the three user classes are removed in sequence, starting from the highest SNR class: after 10 iterations, the power of class 3 users is reduced by 10 dB, after 22 iterations class 2 users are reduced by 10 dB, and, eventually, after 40 iterations all users are removed from the received signal, meaning that each user is decoded as if it were alone on the channel (the ME converges to ≈ 1). Intuitively, we may say that the iterative decoder (under optimized received SNR distribution) performs *implicit* stripping of the different classes of users.

In order to illustrate the behavior of the proposed low-complexity iterative scheme, Fig. 9 (right) shows the evolution of

the residual user symbol variances for the three classes of users versus the decoder iterations, for the proposed low-complexity algorithm. The dash-dotted curves show the evolution of the nominal user symbol variances assumed by the low-complexity detector. The detector switch points, determined by the above algorithm, occur at $\ell_2 = 10$ and $\ell_1 = 26$. In this case, the low-complexity algorithm achieves the same performance as the unconditional MMSE-IC decoder with 55 instead of 43 iterations. In general, a small degradation in achievable performance may be expected, due to the suboptimality of the low-complexity linear detector.

VI. CONCLUDING REMARKS

In this work, we have extended the DE-GA analysis approach of iterative multiuser joint decoding of [14] to the BP algorithm. As a byproduct, we have extended the analysis of [9] of the IO-MUD in the large-system limit to the case of nonuniform symbols prior probabilities. Based on the DE-GA performance characterization of the BP iterative decoder and of

$$\kappa_j^{(\ell)} = \eta_j^{(\ell)} \frac{1 + \frac{1}{L} \sum_{k=1}^{\mathcal{K}_j} \frac{\xi^{(\ell)} \gamma_k v_k^{(\ell_j)}}{\left(1 + \eta_j^{(\ell)} \xi^{(\ell)} \gamma_k v_k^{(\ell_j)}\right)^2}}{1 + \frac{1}{L} \sum_{k=1}^{\mathcal{K}_j} \frac{\gamma_k v_k^{(\ell)}}{\left(1 + \eta_j^{(\ell)} \xi^{(\ell)} \gamma_k v_k^{(\ell_j)}\right)^2} + \frac{1}{L} \sum_{k=\mathcal{K}_j+1}^K \gamma_k v_k^{(\ell)}}$$

$$\frac{1}{\eta_j^{(\ell)}} = 1 + \frac{1}{L} \sum_{k=1}^{\mathcal{K}_j} \frac{\xi^{(\ell)} \gamma_k v_k^{(\ell_j)}}{1 + \eta_j^{(\ell)} \xi^{(\ell)} \gamma_k v_k^{(\ell_j)}} \quad (58)$$

$$\kappa_{j-1}^{(\ell)} = \eta_{j-1}^{(\ell)} \frac{1 + \frac{1}{L} \sum_{k=1}^{\mathcal{K}_{j-1}} \frac{\gamma_k v_k^{(\ell)}}{\left(1 + \eta_{j-1}^{(\ell)} \gamma_k v_k^{(\ell)}\right)^2}}{1 + \frac{1}{L} \sum_{k=1}^{\mathcal{K}_{j-1}} \frac{\gamma_k v_k^{(\ell)}}{\left(1 + \eta_{j-1}^{(\ell)} \gamma_k v_k^{(\ell)}\right)^2} + \frac{1}{L} \sum_{k=\mathcal{K}_{j-1}+1}^K \gamma_k v_k^{(\ell)}}$$

$$\frac{1}{\eta_{j-1}^{(\ell)}} = 1 + \frac{1}{L} \sum_{k=1}^{\mathcal{K}_{j-1}} \frac{\gamma_k v_k^{(\ell)}}{1 + \eta_{j-1}^{(\ell)} \gamma_k v_k^{(\ell)}} \quad (59)$$

its IC-based approximations, we have formulated the problem of optimal user received SNR distribution in terms of simple linear programming. The proposed approximated analysis technique holds for any user code family under mild conditions of regularity that ensure the concentration theorem of [14], and can be easily applied provided that the symbol-error rate versus SNR function characterizing the user code is known (e.g., it can be obtained by simulation or by using EXIT-diagram methods [36]).

In this work, we applied the DE-GA analysis and the related received SNR distribution optimization technique to the case of classical binary convolutional codes, and we computed the achievable spectral efficiency of convolutionally coded CDMA for given user codes and target BER. We showed that by optimizing the SNR distribution very significant gain in terms of spectral efficiency can be achieved, especially for large $(E_b/N_0)_{\text{sys}}$. Interestingly, the simple unconditional MMSE-IC algorithm performs very close to the BP algorithm in terms of achievable spectral efficiency. Driven by this observation and by the fact that the optimal SNR distribution consists of a small number of discrete SNR levels, we provided an approximated version of the unconditional MMSE-IC algorithm achieving complexity comparable to conventional CDMA receivers. As a byproduct of the development of the proposed low-complexity algorithm, we obtained an interesting general expression for the ME of an MMSE filter with mismatched user powers, in the large-system limit. We hasten to say that the *concentrated* nature of the optimized received SNR distribution, for which only a small set of different SNR levels is required, might not hold for other user code families (e.g., when convolutional codes are replaced by LDPCs [43], [44] or turbo codes [45]).

We wish to conclude this work by pointing out some observations about the practical relevance of our findings in CDMA system design.

- For given user codes, decoding algorithm, target BER, and $(E_b/N_0)_{\text{sys}}$, the resulting optimal received SNR distribution can be regarded as the target distribution of SNR that some *power control* algorithm should enforce at the receiver. Notice that standard power control aims at inducing a constant SNR distribution at the receiver, which we have seen to be strongly suboptimal with iterative joint detection.
- In a near-far environment, where users are affected by very different propagation channel gains (due to distance from the base station and to other propagation factors such as fading), in order to induce the optimal received SNR distribution it is convenient to assign the users with largest channel gain to the highest SNR level, and so forth, so that each user can attain its required received SNR level with minimal transmit power. This goes precisely in the opposite direction of conventional power control, that requires that users with the smallest channel gain transmit at the largest power level, in order to render the received SNR of all users constant. Under a reasonable mobility assumption, for which the channel gain of each user is an ergodic process varying on a time scale much larger than the duration of a codeword, the optimal received SNR distribution

will also provide longer battery life to the user terminals with respect to conventional power control.

- For the same reason, in a multicell environment the optimal (single-cell) SNR distribution is expected to provide a smaller total emitted energy from each cell, thus reducing the intercell interference (not taken into account by the iterative joint decoder at each base station). Therefore, the impact of iterative multiuser joint decoding with optimal SNR distribution on the spectral efficiency of a cellular system might be even more evident than in the standard multiple-access (single-cell) scenario examined in this work.

Although very interesting, the issue of a power control algorithm inducing the required received SNR distribution while maximizing the user battery life and/or minimizing the total intercell interference is out of the scope of this work and is left for future investigation.

APPENDIX A PROOF OF PROPOSITION 2

The proof of Proposition 2 follows closely the analysis technique of the IO-MUD developed in [9] for uniform symbol *a priori* probabilities and constant user power, extended in [46] to the case of an arbitrary user power distribution. This technique is based on the *Replica Method*, which is a common tool in statistical mechanics [47]. The main difference between [9] and the case at hand is that here, at any given iteration $\ell > 0$, the IO-MUD (15) for user k treats the messages $\{\mathcal{L}_{j,n}^{\text{dec}} : j \neq k\}$ provided by the SISO decoders at the previous iteration as (log-ratios) prior probabilities for the interfering user symbols. Therefore, Proposition 2 is proved by simply extending the result of [9] to the case of arbitrary symbol prior probabilities under the assumption that, as $K \rightarrow \infty$, the empirical distribution of these prior probabilities converges almost everywhere to some deterministic distribution. Due to the similarity of our proof and [9], we shall give the details of the different steps, while we briefly outline the common parts.

With reference to the channel model (1), in this section we assume, without loss of generality, that

$$\mathbb{E}[w_k^2] = \frac{1}{K} \sum_{k=1}^K w_k^2 = 1.$$

This implies that the average (over the users) received SNR is $\mathbb{E}[\gamma_k] = 1/\sigma_0^2$. The pdf of the channel output \mathbf{y}_n conditioned on the spreading sequences \mathbf{S} is proportional to the quantity

$$Z(\mathbf{y}_n, \mathbf{S}) = \sum_{\mathbf{x}_n \in \{\pm 1\}^K} \Pr(\mathbf{x}_n) \exp\left(-\frac{1}{2\sigma^2} |\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{x}_n|^2\right) \quad (61)$$

when the fictitious noise variance σ^2 is let equal to the true noise variance σ_0^2 .⁹ The distribution of $Z(\mathbf{y}_n, \mathbf{S})$, seen as a function

⁹We keep from [9] the artifact of writing the free energy for an arbitrary σ^2 not necessarily equal to σ_0^2 as it makes the following derivations parallel to [9]. It would be possible to let $\sigma^2 = \sigma_0^2$ from the beginning but, in this case, the saddle-point equations leading to the calculation of the ME would have been of a quite different form from those in [9] and explicit rederivation of several intermediate steps would have been necessary.

of \mathbf{y}_n and \mathbf{S} , is independent of n since the channel input is stationary. Thus, the time index n is dropped and n is used for different purpose in the rest of this section.

In statistical mechanics, the quantity

$$\mathcal{F}_K(\mathbf{y}, \mathbf{S}) = \frac{1}{K} \log Z(\mathbf{y}, \mathbf{S}) \quad (62)$$

is called the *free energy*. One of the basic assumptions adopted in statistical mechanics of disordered systems is that the free energy is self-averaging in the large system limit. That is,

$$\lim_{K \rightarrow \infty} \mathcal{F}_K(\mathbf{y}, \mathbf{S}) = \lim_{K \rightarrow \infty} \mathbb{E}[\mathcal{F}_K(\mathbf{y}, \mathbf{S})] \triangleq \mathcal{F} \quad (63)$$

with probability 1, where averaging is with respect to the random spreading sequences and the channel noise. As shown in [9], \mathcal{F} can be computed via the Replica Method, as the saddle point of a function of some auxiliary variables. The system of equations characterizing this saddle point yields the fixed point (29) for the ME. A fundamental principle in statistical mechanics is that a physical system at the equilibrium maximizes free energy. Hence, if (29) has two stable solutions, the ME is given by the solution which maximizes \mathcal{F} . In passing, we notice that $-\alpha \mathcal{F}|_{\sigma^2=\sigma_0^2} + \frac{1}{2} \log 2\pi\sigma_0^2$ is the large-system differential entropy of the output \mathbf{y} per dimension. The mutual information (in nats) per dimension is given by

$$\begin{aligned} \mathcal{I} &= -\alpha \mathcal{F}|_{\sigma^2=\sigma_0^2} + \frac{1}{2} \log 2\pi\sigma_0^2 - \frac{1}{2} \log 2\pi e\sigma_0^2 \\ &= -\alpha \mathcal{F}|_{\sigma^2=\sigma_0^2} - \frac{1}{2}. \end{aligned} \quad (64)$$

Therefore, maximizing the free energy corresponds to minimizing the mutual information. In the remainder of this section, we shall use the Replica Method (see [9]) in order to determine \mathcal{F} and the fixed point (29).

The Replica Method consists of rewriting the free energy in the following way:

$$\mathcal{F} = \lim_{K \rightarrow \infty} \frac{1}{K} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \log (\mathbb{E}[Z^n(\mathbf{y}, \mathbf{S})]) \quad (65)$$

with the advantage that the expectation operator has moved into the argument of the logarithm. The free energy is evaluated for integer n and the result is assumed to generalize to positive real n . Further discussion about the Replica Method and its justification is provided in [47], [9].

Following [9], by using the *Replica Symmetry* assumption (see [9], [46]), the Gärtner–Ellis theorem, and the Varadhan lemma [48], we get that the free energy can be expressed as follows. Let $\xi(n)$ be defined by the following saddle point:

$$\begin{aligned} \xi(n) = \sup_{m, q} \inf_{\tilde{m}, \tilde{q}} \left\{ \frac{1}{\alpha} G(m, q) - n\tilde{m}m - \frac{n(n-1)}{2} \tilde{q}q \right. \\ \left. + \mathbb{E}_{V, \mathcal{L}} [\log \Phi(V\tilde{m}, V\tilde{q}, \mathcal{L})] \right\} \end{aligned} \quad (66)$$

where n is the replica order. Then

$$\mathcal{F} = \lim_{n \rightarrow 0} \frac{d}{dn} \xi(n). \quad (67)$$

The function $G(m, q)$ in (66) does not depend on the symbols' prior probabilities and is directly obtained from [9] as

$$G(m, q) = \frac{1}{2} \log \frac{(1 + \frac{\alpha}{\sigma^2}(1-q))^{1-n}}{1 + \frac{\alpha}{\sigma^2}(1-q) + \frac{n\sigma_0^2}{\sigma^2}(1 + \frac{\alpha}{\sigma_0^2}(1-2m+q))}. \quad (68)$$

In (66), V, \mathcal{L} are random variables with joint distribution equal to the limit of the empirical joint distribution of

$$V_k \triangleq w_k^2, \quad \mathcal{L}_k \triangleq \log \frac{\Pr(x_k = +1)}{\Pr(x_k = -1)}$$

for $K \rightarrow \infty$, and $\Phi(\tilde{m}, \tilde{q}, \mathcal{L})$ is the moment-generating function of the random vector

$$\{x_a x_b : a = 1, \dots, n, 0 \leq b < a\}$$

for x_i i.i.d., with pmf defined by $\log \frac{\Pr(x_i = +1)}{\Pr(x_i = -1)} = \mathcal{L}$, and computed in the arguments

$$\lambda_{ab} = \begin{cases} \tilde{m}, & b = 0, a \neq b \\ \tilde{q}, & b > 0, a \neq b. \end{cases} \quad (69)$$

The symmetry relation (69) imposed on the moment-generating function variables λ_{ab} reflects the Replica Symmetry assumption, and its validity must be checked *a posteriori*. This can be done by following the same argument as that leading to [9, Proposition 4]. The novelty with respect to [9] is provided by the following lemma, proved at the end of this section.

Lemma 2.1: The moment-generating function $\Phi(\tilde{m}, \tilde{q}, \mathcal{L})$ is given by (70) at the bottom of the page, where $T \triangleq \tanh(\mathcal{L}/2)$. \square

Let $F(n, m, q, \tilde{m}, \tilde{q})$ denote the argument of the extremization (sup-inf) in (66). The saddle-point condition is obtained by the set of equations

$$\frac{\partial}{\partial m} F = 0, \quad \frac{\partial}{\partial q} F = 0, \quad \frac{\partial}{\partial \tilde{m}} F = 0, \quad \frac{\partial}{\partial \tilde{q}} F = 0. \quad (71)$$

In order to obtain \mathcal{F} we should: 1) find the solution of the system of (71); 2) find $\xi(n)$ by substituting this solution into $F(n, m, q, \tilde{m}, \tilde{q})$; 3) evaluate the derivative and the limit in (67). As a matter of fact, from a continuity argument, it is equivalent (but easier) to solve for the system of equations

$$\begin{aligned} \lim_{n \rightarrow 0} \frac{\partial}{\partial m} F = 0, \quad \lim_{n \rightarrow 0} \frac{\partial}{\partial q} F = 0, \quad \lim_{n \rightarrow 0} \frac{\partial}{\partial \tilde{m}} F = 0, \\ \lim_{n \rightarrow 0} \frac{\partial}{\partial \tilde{q}} F = 0 \end{aligned} \quad (72)$$

and substitute the solution into

$$\lim_{n \rightarrow 0} \frac{d}{dn} F(n, m, q, \tilde{m}, \tilde{q}).$$

After some algebra, using the identity (77), the saddle-point equations (72) can be put in the explicit form of (73) at the bottom of the following page. By inspection, we find that the saddle point given by (73) is obtained for $m = q$ and $\tilde{m} = \tilde{q}$. By following the argument leading to [9, Proposition 4], it is possible to show that the Replica Symmetry assumption made

$$\Phi(\tilde{m}, \tilde{q}, \mathcal{L}) = \frac{\int_{\mathbb{R}} \left[(1+T) \cosh^n(z\sqrt{\tilde{q}} + \tilde{m} + \mathcal{L}/2) + (1-T) \cosh^n(z\sqrt{\tilde{q}} + \tilde{m} - \mathcal{L}/2) \right] Dz}{2 \cosh^n(\mathcal{L}/2) \exp(n\tilde{q}/2)} \quad (70)$$

before in order to obtain (66) is indeed valid for the stable solutions of the IO-MUD with arbitrary symbol priors.

Next, we notice that in (15), when computing the output message for user k , only the symbol prior probabilities of interfering users $j \neq k$ are used while the *a priori* pmf for the user k symbol is uniform $\Pr(x_k = 1) = 1/2$. Therefore, the IO-MUD error probability for user k is given by $\Pr(\mathcal{L}_k^{\text{mud}} \leq 0 | x_k = +1)$ even for arbitrary prior probabilities on the other users. It follows from [9] that, in the large-system limit

$$\Pr(\mathcal{L}_k^{\text{mud}} \leq 0 | x_k = +1) = Q\left(\sqrt{\tilde{m}w_k^2}\right)$$

where \tilde{m} is determined by the saddle point (73). Hence, $\sigma_0^2 \tilde{m}$ (independent of the user index k) is the ME of the IO-MUD in the large-system limit with arbitrary prior probabilities on the interfering symbols. By using the first and the third equation in the system (73), replacing q and \tilde{q} by m and \tilde{m} , respectively, and by substituting $\tilde{m} = \Psi/\sigma_0^2$, $1/\sigma_0^2 = \mathbb{E}_\gamma[\gamma]$, we obtain the implicit expression for the ME of the IO-MUD with arbitrary prior

probabilities on the symbols of the interfering users as (74) at the bottom of the page. Finally, by using the fact that \mathcal{L} given γ is $\sim \mathcal{N}(2\mu(\gamma\eta), 4\mu(\gamma\eta))$ and by substituting $T = \tanh(\mathcal{L}/2)$ in (74) we obtain (29). By computing $\lim_{n \rightarrow 0} \frac{\partial}{\partial n} F(n, m, q, \tilde{m}, \tilde{q})$ in the saddle-point solution given by (73), and rewriting the result in terms of Ψ , we obtain the corresponding expression for the free energy \mathcal{F} . As a matter of fact, we prefer to give the result in terms of the mutual information \mathcal{I} that, after some algebra, can be put in the form (30), expressed in bits per dimension.

As anticipated above, standard results in statistical mechanics [9] yield that when (73) has multiple stable solutions, the performance of the IO-MUD corresponds to the solution minimizing the mutual information. This proves all statements in Proposition 2.

Proof of Lemma 2.1: We write explicitly the moment-generating function as (75) and (76) at the bottom of the page, where we have used the identity (Hubbard–Stratonovich transform [9])

$$e^{AX^2/2} = \int_{\mathbb{R}} e^{\pm\sqrt{AX}z} Dz$$

$$\begin{aligned} m &= \mathbb{E}_{V, \mathcal{L}} \left[V \int_{\mathbb{R}} \left\{ \frac{1+T}{2} \tanh(z\sqrt{V\tilde{q}} + V\tilde{m} + \mathcal{L}/2) + \frac{1-T}{2} \tanh(z\sqrt{V\tilde{q}} + V\tilde{m} - \mathcal{L}/2) \right\} Dz \right] \\ q &= \mathbb{E}_{V, \mathcal{L}} \left[V \int_{\mathbb{R}} \left\{ \frac{1+T}{2} \tanh^2(z\sqrt{V\tilde{q}} + V\tilde{m} + \mathcal{L}/2) + \frac{1-T}{2} \tanh^2(z\sqrt{V\tilde{q}} + V\tilde{m} - \mathcal{L}/2) \right\} Dz \right] \\ \tilde{m} &= \frac{1}{\sigma^2 + \alpha(1-q)}, \quad \tilde{q} = \frac{\sigma_0^2 + \alpha(1-2m+q)}{(\sigma^2 + \alpha(1-q))^2}. \end{aligned} \quad (73)$$

$$\begin{aligned} \frac{1}{\Psi} &= 1 + \alpha \left(\frac{1}{\sigma_0^2} - \mathbb{E}_{\gamma, \mathcal{L}} \left[\gamma \int_{\mathbb{R}} \left\{ \frac{1+T}{2} \tanh\left(z\sqrt{\gamma\Psi} + \gamma\Psi + \frac{\mathcal{L}}{2}\right) \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{1-T}{2} \tanh\left(z\sqrt{\gamma\Psi} + \gamma\Psi - \frac{\mathcal{L}}{2}\right) \right\} Dz \right] \right) \\ &= 1 + \alpha \mathbb{E}_{\gamma, \mathcal{L}} \left[\gamma (1-T^2) \int_{\mathbb{R}} \frac{1 - \tanh(z\sqrt{\gamma\Psi} + \gamma\Psi)}{1 - T^2 \tanh^2(z\sqrt{\gamma\Psi} + \gamma\Psi)} Dz \right]. \end{aligned} \quad (74)$$

$$\begin{aligned} \Phi(\tilde{m}, \tilde{q}, \mathcal{L}) &= \sum_{(x_0, x_1, \dots, x_n) \in \{\pm 1\}^{n+1}} \left(\prod_{a=0}^n \Pr(x_a) \right) \exp \left(\sum_{b < a} x_a x_b \lambda_{ab} \right) \\ &= \sum_{(x_0, x_1, \dots, x_n) \in \{\pm 1\}^{n+1}} \left(\prod_{a=0}^n \Pr(x_a) \right) \exp \left[\tilde{m} x_0 \sum_{a=1}^n x_a + \frac{\tilde{q}}{2} \left(\sum_{a=1}^n x_a \right)^2 - \frac{n\tilde{q}}{2} \right] \\ &= \sum_{(x_1, \dots, x_n) \in \{\pm 1\}^{n+1}} \left(\prod_{a=1}^n \Pr(x_a) \right) \left\{ \frac{1+T}{2} \exp \left[\tilde{m} \sum_{a=1}^n x_a + \frac{\tilde{q}}{2} \left(\sum_{a=1}^n x_a \right)^2 \right] \right. \\ &\quad \left. + \frac{1-T}{2} \exp \left[-\tilde{m} \sum_{a=1}^n x_a + \frac{\tilde{q}}{2} \left(\sum_{a=1}^n x_a \right)^2 \right] \right\} e^{-n\tilde{q}/2} \end{aligned} \quad (75)$$

$$\begin{aligned} &= \sum_{(x_1, \dots, x_n) \in \{\pm 1\}^{n+1}} \left(\prod_{a=1}^n \Pr(x_a) \right) \left\{ \frac{1+T}{2} \int_{\mathbb{R}} \exp \left[(z\sqrt{\tilde{q}} + \tilde{m}) \sum_{a=1}^n x_a \right] Dz \right. \\ &\quad \left. + \frac{1-T}{2} \int_{\mathbb{R}} \exp \left[-(z\sqrt{\tilde{q}} + \tilde{m}) \sum_{a=1}^n x_a \right] Dz \right\} e^{-n\tilde{q}/2} \end{aligned} \quad (76)$$

and we have used the definition $T = \tanh(\mathcal{L}/2)$, so that $\Pr(x_0 = +1) = (1 + T)/2$, $\Pr(x_0 = -1) = (1 - T)/2$.

In passing, we notice here that if \mathcal{L}_k is the extrinsic log-ratio provided by the SISO decoder for user k with input SNR $V_k\eta/\sigma_0^2$ then, under the GA, we have that for a given V_k

$$\mathcal{L}_k \sim \mathcal{N}(2\mu(V_k\eta/\sigma_0^2), 4\mu(V_k\eta/\sigma_0^2)).$$

Hence, the limiting joint distribution of V , \mathcal{L} is completely defined by $V/\sigma_0^2 \sim F_\gamma$, and

$$\mathcal{L} \sim \mathcal{N}(2\mu(V\eta/\sigma_0^2), 4\mu(V\eta/\sigma_0^2))$$

given V .

Also, from (75) it is straightforward to show the identity

$$\frac{\partial}{\partial \tilde{q}} \Phi(\tilde{m}, \tilde{q}, \mathcal{L}) = \frac{1}{2} \frac{\partial^2}{\partial \tilde{m}^2} \Phi(\tilde{m}, \tilde{q}, \mathcal{L}) - \frac{n}{2} \Phi(\tilde{m}, \tilde{q}, \mathcal{L}) \quad (77)$$

which is useful to compute the partial derivatives to obtain the saddle point (73).

We can further simplify the expression of the moment-generating function in (76) by noticing that

$$\begin{aligned} & \Pr(x_a = +1)e^{z\sqrt{\tilde{q}+\tilde{m}}} + \Pr(x_a = -1)e^{-(z\sqrt{\tilde{q}+\tilde{m}})} \\ &= \frac{1+T}{2}e^{z\sqrt{\tilde{q}+\tilde{m}}} + \frac{1-T}{2}e^{-z\sqrt{\tilde{q}+\tilde{m}}} \\ &= \cosh(z\sqrt{\tilde{q}+\tilde{m}}) \left(1 + \tanh(\mathcal{L}/2) \tanh(z\sqrt{\tilde{q}+\tilde{m}})\right) \\ &= \frac{\cosh(z\sqrt{\tilde{q}+\tilde{m}} + \mathcal{L}/2)}{\cosh(\mathcal{L}/2)} \end{aligned}$$

and similarly

$$\begin{aligned} & \Pr(x_a = +1)e^{-(z\sqrt{\tilde{q}+\tilde{m}})} + \Pr(x_a = -1)e^{z\sqrt{\tilde{q}+\tilde{m}}} \\ &= \frac{\cosh(z\sqrt{\tilde{q}+\tilde{m}} - \mathcal{L}/2)}{\cosh(\mathcal{L}/2)}. \end{aligned}$$

Then, by rearranging the terms in the sum with respect to x_1, \dots, x_n in (76) and summing by using the two above identities, we obtain (70).

APPENDIX B

PROOF OF PROPOSITIONS 4 AND 5

For notational simplicity, we drop the time index n . The observed signal after soft IC for user k is given by $\mathbf{y}_k = \mathbf{y} - \bar{\mathbf{y}}_k$, that can be written explicitly as

$$\begin{aligned} \mathbf{y}_k &= \sum_{j \neq k} \mathbf{s}_j w_j (x_j - t_j) + \mathbf{s}_k w_k x_k + \boldsymbol{\nu} \\ &= \sum_{j \neq k} \mathbf{s}_j w_j \tilde{x}_j + \mathbf{s}_k w_k x_k + \boldsymbol{\nu} \end{aligned} \quad (78)$$

where we let $t_j = \tanh(\mathcal{L}_j^{\text{dec}}/2)$ and $\tilde{x}_j = x_j - t_j$. Under the independence assumption, the symbols \tilde{x}_j are still mutually independent and have mean zero. Furthermore, we notice that, under the GA, $\mathcal{L}_j^{\text{dec}} \sim \mathcal{N}(2\mu(\gamma_j\eta)x_j, 4\mu(\gamma_j\eta))$, since the j th decoder input SINR is equal to $\gamma_j\eta$ by assumption. Therefore,

the variance of the symbols \tilde{x}_j for any user j received at power level $\gamma_j = \gamma$ is given by

$$\begin{aligned} \mathbb{E}[|\tilde{x}_j|^2 | \gamma_j = \gamma] &= \mathbb{E}[\mathbb{E}[|x_j - t_j|^2 | \mathcal{L}_j^{\text{dec}}] | \gamma_j = \gamma] \\ &= \mathbb{E}[1 - 2\mathbb{E}[x_j | \mathcal{L}_j^{\text{dec}}] t_j + t_j^2 | \gamma_j = \gamma] \\ &= 1 - \mathbb{E}[t_j^2 | \gamma_j = \gamma] \\ &= 1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)) Dy \end{aligned} \quad (79)$$

where we have used the fact that t_j is a function of $\mathcal{L}_j^{\text{dec}}$ and that, by definition, $t_j = \mathbb{E}[x_j | \mathcal{L}_j^{\text{dec}}]$.

Now, for each $j \neq k$, we can normalize the symbol \tilde{x}_j so that it has unit variance, and replace the amplitude w_j in (78) by the modified amplitude

$$\tilde{w}_j = w_j \sqrt{1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(\gamma_j\eta)} + \mu(\gamma_j\eta)) Dy}.$$

Hence, it is readily seen that the unconditional MMSE filter coincides with the standard MMSE filter for the resulting CDMA system. By applying the large-system analysis of [6] we obtain immediately that the ME under linear MMSE detection is given by (33), and the ME under SUMF is given by (34). This proves Propositions 4 and 5.

APPENDIX C

PROOFS OF PROPOSITION 6 AND COROLLARY 6.1

Fix user k as the reference user. The nominal and the actual interference-plus-noise covariance matrices are given by

$$\begin{aligned} \boldsymbol{\Sigma}_U &= \mathbf{I} + \sum_{j \neq k} U_j \mathbf{s}_j \mathbf{s}_j^T \\ \boldsymbol{\Sigma}_P &= \mathbf{I} + \sum_{j \neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \end{aligned} \quad (80)$$

respectively. The output SINR of the mismatched MMSE filter for user k that assumes $\boldsymbol{\Sigma}_U$ instead of $\boldsymbol{\Sigma}_P$ is given by

$$\beta_k = P_k \frac{(\mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k)^2}{\mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \boldsymbol{\Sigma}_P \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k}. \quad (81)$$

From the standard result of [6] we have that, under the assumptions of Proposition 6, the nominal ME $\eta = \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k$ converges with probability 1 to the unique fixed point of the Tse–Hanly equation (57).

We rewrite the denominator of (81) in the large-system limit as

$$\begin{aligned} \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \boldsymbol{\Sigma}_P \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k &= \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \left(\sum_{j \neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \right) \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k \\ &\quad + \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-2} \mathbf{s}_k \\ &\xrightarrow{\text{w.p.1}} \lim_{K \rightarrow \infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_U^{-1} \left(\sum_{j \neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \right) \boldsymbol{\Sigma}_U^{-1} \right) \\ &\quad + \lim_{K \rightarrow \infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_U^{-2} \right) \end{aligned} \quad (82)$$

where we used the result proved in [49]

$$\lim_{K \rightarrow \infty} \mathbf{s}_k^T \mathbf{A}_K \mathbf{s}_k \stackrel{\text{w.p.1}}{=} \lim_{K \rightarrow \infty} \frac{1}{L} \text{tr}(\mathbf{A}_K) \quad (83)$$

for a sequence of $L \times L$ random matrices \mathbf{A}_K statistically independent of \mathbf{s}_k with well-defined limiting eigenvalue distribution and finite maximum eigenvalue. Notice that for all inverse covariance matrices and powers thereof involved in (82), the maximum eigenvalue is always upper-bounded by 1.

Next, we evaluate the two limits in (82) separately. For the first we have (84) at the bottom of the page, where (a) follows from the matrix inversion lemma, by writing $\boldsymbol{\Sigma}_U = U_j \mathbf{s}_j \mathbf{s}_j^T + \boldsymbol{\Sigma}_{U,j}$, where

$$\boldsymbol{\Sigma}_{U,j} = \mathbf{I} + \sum_{\ell \neq k, \ell \neq j} U_\ell \mathbf{s}_\ell \mathbf{s}_\ell^T.$$

We used repeatedly the lemma (83), and we defined the limits

$$\begin{aligned} A &= \lim_{K \rightarrow \infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_{U,j}^{-1} \right) \\ B &= \lim_{K \rightarrow \infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_{U,j}^{-2} \right). \end{aligned} \quad (85)$$

Under the assumptions of Proposition 6, the matrices $\boldsymbol{\Sigma}_U$, $\boldsymbol{\Sigma}_{U,j}$, and $\boldsymbol{\Sigma}_P$ have all a well-defined limiting eigenvalue distribution and are invertible with probability 1, therefore, the limits in (85) exist and are immediately obtained from the limiting eigenvalue distributions.

For the second limit in (82) we have, again using lemma (83)

$$\lim_{K \rightarrow \infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_U^{-2} \right) = B \quad (86)$$

since the limiting eigenvalue distribution of $\boldsymbol{\Sigma}_U$ and $\boldsymbol{\Sigma}_{U,j}$ coincide (notice that the two matrices differ by the rank-1 matrix $U_j \mathbf{s}_j \mathbf{s}_j^T$; that has no effect on the limiting eigenvalue distribution).

Let $G(\lambda)$ denote the limiting eigenvalue distribution of $\sum_j U_j \mathbf{s}_j \mathbf{s}_j^T$. Then, we can write

$$A = \int \frac{1}{1+\lambda} dG(\lambda)$$

$$B = \int \frac{1}{(1+\lambda)^2} dG(\lambda). \quad (87)$$

Eventually, the last line of (82) is given by

$$D = \alpha \mathbb{E}_{U,P} \left[\frac{P \int \frac{1}{(1+\lambda)^2} dG(\lambda)}{\left(1 + U \int \frac{1}{1+\lambda} dG(\lambda)\right)^2} \right] + \int \frac{1}{(1+\lambda)^2} dG(\lambda). \quad (88)$$

Now, we use the result by Silverstein and Bai (see [6] and references therein) yielding $G(\lambda)$ in terms of its Stieltjes transform

$$m_G(z) \triangleq \int \frac{1}{\lambda - z} dG(\lambda), \quad \text{Im}\{z\} > 0.$$

We have

$$m_G(z) = \frac{1}{-z + \alpha \mathbb{E}_U \left[\frac{U}{1 + U m_G(z)} \right]}. \quad (89)$$

In particular, we have that η defined in (57) is given by

$$\eta = \int \frac{1}{1+\lambda} dG(\lambda) = m_G(-1). \quad (90)$$

Furthermore

$$\int \frac{1}{(1+\lambda)^2} dG(\lambda) = \frac{d}{dz} m_G(z) \Big|_{z=-1} \triangleq m'_G(-1). \quad (91)$$

By substituting (90) and (91) in the SINR denominator (88), and using the result in the expression for the SINR (81) we can write the large-system ME of user k , given by β_k/P_k , as

$$\kappa = \frac{\eta^2}{m'_G(-1) \left(1 + \alpha \mathbb{E}_{U,P} \left[\frac{P}{(1+\eta U)^2} \right]\right)}. \quad (92)$$

Since for the matched MMSE receiver (i.e., for $U_j = P_j, j = 1, \dots, K$), it must be $\eta = \kappa$, we obtain

$$m'_G(-1) = \frac{\eta}{1 + \alpha \mathbb{E}_U \left[\frac{U}{(1+\eta U)^2} \right]}.$$

$$\begin{aligned} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_U^{-1} \left(\sum_{j \neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \right) \boldsymbol{\Sigma}_U^{-1} \right) &= \frac{1}{L} \sum_{j \neq k} P_j \mathbf{s}_j^T \boldsymbol{\Sigma}_U^{-2} \mathbf{s}_j \\ &\stackrel{(a)}{=} \frac{1}{L} \sum_{j \neq k} P_j \mathbf{s}_j^T \left[\boldsymbol{\Sigma}_{U,j}^{-1} - \frac{U_j}{1 + U_j \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-1} \mathbf{s}_j} \boldsymbol{\Sigma}_{U,j}^{-1} \mathbf{s}_j \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-1} \right]^2 \mathbf{s}_j \\ &= \frac{1}{L} \sum_{j \neq k} P_j \left[\mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-2} \mathbf{s}_j + \left(\frac{U_j \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-1} \mathbf{s}_j}{1 + U_j \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-1} \mathbf{s}_j} \right)^2 \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-2} \mathbf{s}_j - \left(\frac{2U_j \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-1} \mathbf{s}_j}{1 + U_j \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-1} \mathbf{s}_j} \right) \mathbf{s}_j^T \boldsymbol{\Sigma}_{U,j}^{-2} \mathbf{s}_j \right] \\ &\xrightarrow{\text{w.p.1}} \lim_{K \rightarrow \infty} \frac{\alpha}{K} \sum_{j \neq k} \frac{P_j B}{(1 + U_j A)^2} \\ &= \alpha \mathbb{E}_{U,P} \left[\frac{PB}{(1 + UA)^2} \right] \end{aligned} \quad (84)$$

By using the above expression in (92) we obtain (56). This concludes the proof of Proposition 6.

Next, we focus on the proof of Corollary 6.1. Let $\{U_k\}$ be a nominal power assignment adapted to the true powers $\{P_k\}$, in the sense that $\sum_{k=1}^K U_k = \sum_{k=1}^K P_k$ and that the sequences $\{P_k\}$ and $\{U_k\}$ are sorted in nondecreasing order by the same permutation. Without loss of generality, we assume that $\{P_k\}$ and $\{U_k\}$ are nondecreasing (i.e., the common sorting permutation is the identity). The nominal powers are a conservative choice if $\kappa \geq \eta$, where κ and η are the true and the nominal ME given by Proposition 6, respectively. From (56), we see that this is verified if and only if

$$\mathbb{E}_U \left[\frac{U}{(1 + \eta U)^2} \right] \geq \mathbb{E}_{P,U} \left[\frac{P}{(1 + \eta U)^2} \right]. \quad (93)$$

Assume, for the time being, a finite number of users. Then, inequality (93) becomes

$$\frac{1}{K} \sum_{k=1}^K \frac{U_k}{(1 + \eta U_k)^2} \geq \frac{1}{K} \sum_{k=1}^K \frac{P_k}{(1 + \eta U_k)^2}. \quad (94)$$

We make use of the following lemma, proved in [50].

Lemma: Let \mathbf{a} , \mathbf{b} , and \mathbf{c} be real vectors of dimension K with nondecreasing components. If \mathbf{b} majorizes \mathbf{a} , Then

$$\sum_{k=1}^K \frac{a_k}{c_k} \geq \sum_{k=1}^K \frac{b_k}{c_k}. \quad (95)$$

□

We apply the lemma to (94) by letting $a_k = U_k$, $b_k = P_k$, and $c_k = (1 + \eta U_k)^2$ and we conclude that if $\{P_k\}$ majorizes $\{U_k\}$, then (94) is verified. Finally, a standard continuity argument extends the result to $K \rightarrow \infty$.

REFERENCES

- [1] S. Verdú, *Multuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [2] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [3] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [4] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.
- [5] —, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1302–1327, May 2001.
- [6] D. N. C. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and capacity," *IEEE Trans. Inform. Theory*, vol. 45, pp. 641–675, Mar. 1999.
- [7] D. N. C. Tse and S. Verdú, "Optimum asymptotic multiuser efficiency of randomly spread CDMA," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2718–2722, Nov. 2000.
- [8] E. Chong, J. Zhang, and D. N. C. Tse, "Output MAI distribution of linear MMSE multiuser receivers in DS-CDMA systems," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1128–1144, Mar. 2001.
- [9] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.
- [10] R. Müller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1429–1441, Aug. 2001.
- [11] G. Caire and R. Müller, "The optimal received power distribution for IC-based iterative multiuser joint decoders," in *Proc. 39th Allerton Conf. Communications, Control and Computing*, Monticello, IL, Oct. 2001.
- [12] R. Müller and G. Caire, "Efficient implementation of iterative multiuser decoding," in *Proc. IEEE Int. Symp. Information Theory (ISIT 2002)*, Lausanne, Switzerland, July 2002, p. 446.
- [13] R. Müller, G. Caire, and T. Tanaka, "Density evolution and power profile optimization for iterative multiuser decoders based on individually optimum multiuser detectors," in *Proc. 40th Allerton Conf. Communications, Control and Computing*, Monticello, IL, Oct. 2002.
- [14] J. Boutros and G. Caire, "Iterative multiuser decoding: Unified framework and asymptotic performance analysis," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1772–1793, July 2002.
- [15] M. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple access channel," in *Proc. Asilomar Conf.*, Pacific Grove, CA, Nov. 1997.
- [16] A. Roumy, S. Guemghar, G. Caire, and S. Verdú, "Maximizing the spectral efficiency of LDPC-encoded CDMA," *IEEE Trans. Inform. Theory*, submitted for publication.
- [17] R. Müller and W. Gerstacker, "On the capacity loss due to separation of detection and decoding," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1769–1778, Aug. 2004.
- [18] "Special Issue on Iterative Decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 493–853, Feb. 2001.
- [19] X. Wang and V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1047–1061, July 1999.
- [20] P. Alexander, A. Grant, and M. Reed, "Iterative detection in code-division multiple-access with error control coding," *Europ. Trans. Telecommun.*, vol. 9, no. 5, pp. 419–425, Sept. 1999.
- [21] H. ElGamal and E. Geraniotis, "Iterative multiuser detection for coded CDMA signals in AWGN and fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 30–41, Jan. 2000.
- [22] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.
- [23] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.
- [24] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.
- [25] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.
- [26] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1639–1542, July 1999.
- [27] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple antenna channel," *IEEE Trans. Commun.*, vol. 51, pp. 389–399, Mar. 2003.
- [28] V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [29] D. Divsalar, M. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 46, pp. 258–268, Feb. 1998.
- [30] T. Richardson, R. Urbanke, and A. Shokrollahi, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.
- [31] H. El Gamal and R. Hammons, "Analyzing the turbo decoder using the Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 671–686, Feb. 2001.
- [32] S. Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.
- [33] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [34] Y. Kuznetsov, *Elements of Applied Bifurcation Theory*. New York: Springer-Verlag, 1998.
- [35] A. Roumy, S. Guemghar, G. Caire, and S. Verdú, "Design methods for irregular repeat-accumulate codes," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1711–1727, Aug. 2004.

- [36] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.
- [37] M. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1928–1946, July 2001.
- [38] L. Cottatellucci and R. R. Müller, "Multiuser interference mitigation with multistage detectors: Design and analysis for unequal powers," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002.
- [39] P. Loubaton and W. Hachem, "Asymptotic analysis of reduced rank Wiener filters," in *Proc. IEEE Information Theory Workshop*, Paris, France, Apr. 2003.
- [40] L. Li, A. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Commun. Inform. Syst.*, vol. 1, pp. 273–304, Fall 2001.
- [41] A. Tarable, G. Montrosi, and S. Benedetto, "A linear front-end for iterative soft interference cancellation and decoding in coded CDMA," in *Proc. IEEE Int. Conf. Communications (ICC)*, Helsinki, Finland, June 2001.
- [42] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. San Diego, CA: Academic, 1979.
- [43] A. Roumy, S. Guemghar, G. Caire, and S. Verdú, "Maximizing the spectral efficiency of coded CDMA under successive decoding," *IEEE Trans. Inform. Theory*, vol. 50, pp. 152–164, Jan. 2004.
- [44] G. Yue and X. Wang, "Coding-spreading optimization for turbo multiuser detection in LDPC-coded CDMA," in *Proc. 3rd Int. Symp. Turbo Codes and Related Topics*, Brest, France, Sept. 2003.
- [45] N. Chayat and S. Shamai, "Convergence properties of iterative soft onion peeling," in *Proc. Information Theory Workshop 1999*, Kruger National Park, South Africa, June 1999, p. 9.
- [46] D. Guo and S. Verdú, "Multiuser detection and statistical mechanics," in *Ian Blake Festschrift*, 2002.
- [47] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*. Oxford, U.K: Oxford Univ. Press, 2001.
- [48] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag, New York, 1998.
- [49] V. Marcenko and L. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. U.S.S.R., Sbornik*, no. 1, pp. 457–483, 1967.
- [50] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1165–1175, July 2004.
- [51] P. Schramm and R. R. Müller, "Spectral efficiency of CDMA systems with Linear MMSE interference suppression," *IEEE Trans. Commun.*, vol. 47, pp. 722–731, May 1999.