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

Giuseppe Caire⋆, Ralf Müller† and Toshiyuki Tanaka‡

March 12, 2003

⋆ Institut Eurecom, 2229 Route des Crétes, B.P. 193, 06904 Sophia-Antipolis, France.
Email: caire@eurecom.fr

† Forschungszentrum Telekommunikation Wien FTW, Tech Gate Vienna, Donau-City Str. 1/3, 1220 Wien, Austria.
Email: mueller@ftw.at

‡ Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan.
Email: tanaka@eei.metro-u.ac.jp

Note: the content of this paper has been partially presented in the 39th Allerton Conference, 2001, in the Int. Symp. on Inform. Theory, ISIT 2002, and in the 40th Allerton Conference, 2002.

[1]

[1] T. Tanaka acknowledges support from EPSRC research grant GR/N00562 and from Grant-in-aid for scientific research on priority areas 14084209, MEXT, Japan.

Abstract

We consider a canonical model for coded 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 exact BP decoder and of some suboptimal approximations based on *Interference Cancellation* (IC). Based on this analysis, we optimize the received user 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. 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 exact 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 of power-mismatched MMSE filters in the large-system limit, and we extend the analysis of the symbol-by-symbol MAP multiuser detector in the large-system limit to the case of non-constant user powers and non-uniform symbol prior probabilities.

Keywords: Multiuser Detection, Multiple-Access Channel Capacity, Iterative Decoding, Statistical Mechanics.

1 Problem statement and prior work

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

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

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 i.i.d., with variance per component σ_0^2 (we write $\boldsymbol{\nu}_n \sim \mathcal{N}(\mathbf{0}, \sigma_0^2\mathbf{I})$).

As usual, in multiple-access channels the users send independent and independently encoded information [2] (see the block-diagram in Fig. 1). This implies that $\boldsymbol{\Sigma}_x = E[\mathbf{x}_i\mathbf{x}_i^T]$ is diagonal. Without loss of generality, we let $\boldsymbol{\Sigma}_x = \mathbf{I}$ and 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) of user k. We let $\mathcal{C}_k$ denote the user codebooks, of rate $R_k = \frac{1}{N}\log_2|\mathcal{C}_k|$ bit per symbol. 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 $(\widehat{m}_1, \dots, \widehat{m}_K)$. 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

$$P_b^{(k)} = \frac{1}{B_k}\sum_{j=1}^{B_k} \Pr\left(\widehat{b}_{k,j} \neq b_{k,j}\right) \tag{2}$$

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, where T is the (continuous-time) duration of one channel use. Therefore, the system spectral efficiency is given by [4]

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

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]

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

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). Moreover, 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., the entries of $\mathbf{S}$ are generated i.i.d. according to some probability distribution) are derived based on this model for the sake of analytical tractability (see for example [4, 6, 7, 8, 9, 10, 11, 12, 13]). We shall not question here the validity of this widely accepted model. Nevertheless, we would like to stress the fact that both more refined analysis and practical experience shows that the conclusions drawn from the real canonical model (1) apply (at least qualitatively) to more complicated and close-to-practice models taking into account complex-valued baseband equivalent channels [5], asynchronous transmission [14] and transmission through multipath fading channels [15] with imperfect channel estimation. The main fact that makes the model (1) "close" to practical CDMA settings is the random spreading assumption, which prevents the users to pick their waveforms optimally. In this respect, the random-spreading point of view reflects real-life CDMA practice [16], where physical impairments and practical constraints prevent the system from optimizing the user waveforms.

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 $\{\mathcal{C}_k : k = 1, \ldots, K\}$, given iterative joint decoders (see [17] and references therein) and subject to the individual maximum BER constraints $P_b^{(k)} \leq \epsilon$ for all $k = 1, \ldots, K$, 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 CDMA and by providing a preview of the reminder 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., $\epsilon \to 0$) [2] was found by Verdú and Shamai in [4, 5] for given finite *channel load* $K/L = \alpha$, and reads

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

where η is the solution to [6]

$$\frac{1}{\eta} = 1 + \alpha\frac{\gamma}{1+\eta\gamma} \tag{6}$$

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

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

[2]The 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].

The spectral efficiency (5) is achieved by Gaussian codebooks and constant user received SNR.[3] The supremum of C over $\alpha \geq 0$ is obtained for $\alpha \to \infty$ (an infinite number of users per dimension, with vanishing user coding rate), and is given by the single-user Gaussian channel spectral efficiency,

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \frac{2^{2\mathsf{C}} - 1}{2\mathsf{C}} \tag{7}$$

It is interesting to notice that, in order to achieve (5), an optimal (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 $k = 1, 2, \ldots, K$) 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 [18]. 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, LDPC codes) for successive stripping decoding is studied in [19].

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

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

where η is the solution to [9]

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

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

It is not hard to show that, for given γ, the maximum of C in (8) is also obtained by letting $\alpha \to \infty$ 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 $k = 1, \ldots, K$, followed by a

[3] 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.

bank of K single-user decoders, each processing its own MUD front-end output and producing the decoded message $\widehat{m}_k$ 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 spectral efficiency achieved by *linear* MUD (single-user matched filter (SUMF), linear 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 MAP symbol-by-symbol estimation for Gaussian signals. In [20], Müller and Gerstacker found the spectral efficiency with binary user codes and the individually optimal (symbol-by-symbol MAP) MUD front-end. Remarkably, both for Gaussian and for 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, 20]

$$\mathsf{C}^{\rm sep} = \mathsf{C}^{\rm joint} + \frac{1}{2}\log_2 \eta + \frac{1-\eta}{2}\log_2 e \tag{10}$$

where $\mathsf{C}^{\rm joint}$ 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 coding (see [21] and references therein), "Turbo" multiuser joint decoding was proposed in several works (see for example [22, 23, 24] and references in [17]). These algorithms seek a trade-off 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 3) was provided by Boutros and Caire in [17]. This analysis is based on the general technique known as *Density Evolution* (DE) [25], commonly used to determine the iterative decoding limits of Turbo Codes and LDPC codes, and is exact in the limit of large blocklength (notice: to obtain a meaningful large system limit we let *first* $N \to \infty$ and *then* $K \to \infty$ with $K/L = \alpha$).

Preview of this paper. Several issues are left open in [17]. In particular, how the *exact* BP decoder compares with respect to its IC-based approximations? What is the optimal received SNR distribution maximizing 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 (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 outperform the separated approach?

In this work, we provide answers to the above questions. In Section 2 we recall the exact BP decoder and some lower-complexity approximations based on IC. In Section 3 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 4 we provide a simple linear programming algorithm for the optimization of the received SNR distribution. Our results show that, under constant received SNR, the exact 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 exact BP 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 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 5 we provide a low-complexity approximated version of the MMSE-IC iterative decoder that offers a very competitive trade-off between complexity and performance. Finally, we point out our conclusions in Section 6. The proofs of the main results are provided in the Appendix.

2 Iterative joint decoding algorithm

In the rest 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

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

The BP algorithm [26, 27] approximates iteratively the log-ratios $\mathcal{L}_{k,j}^{\rm bit}$ corresponding to the marginals of the a posteriori joint pmf $\Pr(\mathbf{b}_1, \ldots, \mathbf{b}_K|\mathbf{Y})$ of the user information bits. After a given number of iterations, a symbol-by-symbol decision is made according to the threshold rule

$$\widehat{b}_{k,j} = \mathrm{sign}(\mathcal{L}_{k,j}^{\rm bit}) \tag{12}$$

Standard results [27] show that if the *dependency graph* describing $\Pr(\mathbf{b}_1, \ldots, \mathbf{b}_K|\mathbf{Y})$ is cycle-free, then BP yields symbol-by-symbol 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 as long as $K > 1$ and the user codes are non-trivial (i.e., have rate $R_k < 1$). Nevertheless, for sufficiently large blocklength 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 girth ℓ decreases linearly with N [17]. 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 [25]. 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 MAP multiuser detector (IO-MUD) (see the block-diagram in Fig. 2).

SISO decoding is formally expressed by

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

for all $k = 1, \ldots, K$ and $n = 1, \ldots, N$, where $\mathcal{L}^{\rm mud}_{k,j}$ is the message (log-ratio) sent by the IO-MUD for user k relative to coded symbol $c_{k,j}$ and $\mathcal{L}^{\rm dec}_{k,n}$ is the so called decoder "extrinsic information". For convolutional codes, (13) is efficiently implemented by the well-known forward-backward algorithm [28]. The same forward-backward algorithm can compute the log-ratios $\{\mathcal{L}^{\rm bit}_{k,j} : j = 1, \ldots, B_k\}$ for the user information bits while computing (13).

IO-MUD consists of calculating the a posteriori log-ratios

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

$$= \log \frac{\sum_{\mathbf{x}\in\{\pm1\}^K : x_k=+1} \exp\left(-\frac{1}{2\sigma_0^2}\left|\mathbf{y}_n - \sum_{j=1}^K w_j \mathbf{s}_j x_j\right|^2 + \frac{1}{2}\sum_{j\neq k} x_j \mathcal{L}^{\rm dec}_{j,n}\right)}{\sum_{\mathbf{x}\in\{\pm1\}^K : x_k=-1} \exp\left(-\frac{1}{2\sigma_0^2}\left|\mathbf{y}_n - \sum_{j=1}^K w_j \mathbf{s}_j x_j\right|^2 + \frac{1}{2}\sum_{j\neq k} x_j \mathcal{L}^{\rm dec}_{j,n}\right)} \tag{15}$$

for all $k = 1, \ldots, K$ and $n = 1, \ldots, N$. Unfortunately, there is no efficient way to perform this calculation, in general. [4]

[4]Based on the fact that $\mathbf{SWx}$, 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

Various schemes have been proposed to simplify the exact 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 $\{\mathcal{L}^{\rm dec}_{j,n} : j \neq k\}$, given by

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

where the filter $\mathbf{h}_{k,n}$ minimizes the *conditional* MSE

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

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

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

where

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

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

$$\mathcal{L}^{\rm mud}_{k,n} = 2\beta_{k,n} z_{k,n} \tag{19}$$

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.

Decoder [29]) 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 [30]. Nevertheless, this approach is prohibitively complex for large K and/or $\alpha > 1$ (i.e., $K > N$).

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

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

Then, the linear MMSE estimation of the zero-mean symbol $x_{k,n}$ is obtained by filtering the zero-mean observation $\mathbf{y}_n - \bar{\mathbf{y}}_n$. This decomposition of linear MMSE estimators is canonical [31]. However, it is interesting to notice that, in this setting, the elimination of the conditional mean of the observation takes on the meaning of *soft* Interference Cancellation (IC). In fact, $\bar{\mathbf{y}}_n$ is the (non-linear) 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 $\{\mathcal{L}^{\text{dec}}_{j,n} : j \neq k\}$.

Since (16) is obtained by solving a 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 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 MSE

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

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

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

where

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

is the output signal to interference plus noise ratio (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 $E[\tanh^2(\mathcal{L}^{\text{dec}}_{j,n}/2)]$ can be replaced by the empirical mean

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

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

The unconditional MMSE-IC scheme requires the evaluation of only one filter per user per iteration.

Single-user matched filter with IC. A further simplfication is obtained by replacing the MMSE filter by the single-user matched filter (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 [32]), and 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.

3 Density evolution 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 girth ℓ vanishes as the blocklength N increases), a general concentration theorem [25] 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 $N \to \infty$. In [17] 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 convolutional codes with random independent interleaving.

In the rest of this paper we make the following assumptions: 1) the user codes are all derived by the same convolutional code $\mathcal{C}$ of rate R, and differ only by the interleaver randomly and independently generated for each user; 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 (non-random) distribution $F_\gamma(z)$, as $K \to \infty$; 4) as anticipated before, we shall study the large-system limit of the iterative decoders by letting first

$N \to \infty$ (to approach the concentration theorem limit) and then $K \to \infty$ 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, uncoditional 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 blocklength N) are asymptotically conditionally independent given the k-th user transmitted codeword.

Proof. It follows directly as a corollary of [8, 9, 13]. □

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 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 defined above. The pdf of the log-ratio $\mathcal{L}_{k,n}^{\text{dec}}$ defined in (13) satisfies the symmetry condition [33]

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

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*

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

This is equivalent to model $\mathcal{L}_{k,n}^{\text{dec}}$ 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 $\delta_{k,n} \sim \mathcal{N}(0, 1/\mu_k^{(\ell)})$. The above *Gaussian Approximation* has been used extensively to study the performance of Turbo Codes [34] and LDPC codes [35] 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 [17]. 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

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

Hence, we let

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

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 Gaussian Approximation). Eventually, the DE with Gaussian Approximation (referred to as DE-GA in the following) will take on the form of the one-dimensional dynamical system

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

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 exact BP decoder is given by the stable solution to the fixed-point equation

$$\frac{1}{\Psi} = 1 + \alpha E_\gamma \left[\gamma \int_{\mathbb{R}^2} \frac{\left(1 - \tanh^2\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right)\right)\left(1 - \tanh\left(z\sqrt{\gamma\Psi} + \gamma\Psi\right)\right)}{1 - \tanh^2\left(y\sqrt{\mu(\gamma\eta)} + \mu(\gamma\eta)\right)\tanh^2\left(z\sqrt{\gamma\Psi} + \gamma\Psi\right)} Dz Dy\right] \tag{29}$$

in the interval $[0,1]$ that minimizes the quantity

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

where $E_\gamma[\cdot]$ denotes expectation with respect to the received SNR distribution F_γ and where, from (27), we define the function

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

Proof. See Appendix A. □

Equation (29) may have either one (see example in Fig. 3, left) or three distinct solutions (see example in Fig. 3, right) in the interval $[0,1]$, depending on η, α and F_γ. If (29) has three solutions $0 \le \Psi_1 < \Psi_2 < \Psi_3 \le 1$, 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 bit/s/Hz) for the channel (1) where the input symbols $\mathbf{x}_n$ are binary with non-uniform a priori marginal pmf given by

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

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

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

converges almost everywhere as $K \to \infty$ 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 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

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

in the interval $[0, 1]$, where $\mu(z)$ is defined in (31).

Proof. See [17]. □

Proposition 4. The mapping function $\Psi(\eta)$ for the unconditional MMSE-IC decoder is given by the *unique* solution to

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

in the interval $[0, 1]$, where $\mu(z)$ is defined in (31).

Proof. See [11]. □

Although not surprising, it is interesting to notice that equations (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

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

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

Proof. See [11]. □

Proposition 6. 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.

Proof. It follows directly from the definition. □

With some effort, it is possible to verify that, for all $\eta \in [0,1]$ and all SNRs distributions F_γ the following inequalities hold

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

Moreover, for any finite α we have $\eta^{(0)} > 0$ and the functions Ψ are non-decreasing 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 $\eta^\star$ denote this solution, we have $\lim_{\ell\to\infty} \eta^{(\ell)} = \eta^\star$.

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 $\eta^\star$. In fact, after many iterations, every k-th SISO decoder "sees" a binary-input AWGN channel with SNR $\gamma_k\eta^\star$. Therefore, for a given user code $\mathcal{C}$, the BER is uniquely determined by $\eta^\star$ and by the individual received SNR γ_k. For example, if $\eta^\star \approx 1$, every user in the system attains a performance close to its single-user lower bound, as if it was 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 code with (octal notation [36]) generators $(133,171)_8$. Figs. 4 and 5 show BER vs. 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 [17], this transition corresponds to a *fold bifurcation* [37] of the dynamical system (28) representing the DE-GA. The value of α for which the bifurcation appears depends on the decoder algorithm. For example, for $\alpha = 1.0$ (Fig. 4) the SUMF-IC decoder shows the bifurcation behavior while the other detectors have a smooth waterfall. For $\alpha = 2.0$ (Fig. 5) the exact 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).

4 Received SNR distribution optimization

In this section we aim at optimizing the received SNR distribution F_γ in order to maximize the spectral efficiency $\rho = \alpha R$, for a given user convolutional code $\mathcal{C}$, given channel load α, given maximum BER constraint ϵ (for all users in the system) and for a given iterative decoder in the class of algorithms studied in the previous sections.

For simplicity, we quantize the SNRs levels, i.e., we shall assume that the users received SNRs take on values in a finite discrete set of levels $0 < g_1 < g_2 < \ldots < g_J$, for some finite integer J. Users received at SNR level g_j are said to "belong to class j". Moreover, we define the partial channel loads $\alpha_j = K_j/L$, for $j = 1, \ldots, J$, where K_j is the number of users in class j. Clearly, $\sum_{j=1}^J \alpha_j = \alpha$. Finally, we assume that when $K \to \infty$ all the class sizes K_j grow to infinity, with given ratios $K_j/K = \alpha_j/\alpha$. In order to stress the dependency of the DE-GA mapping function on the system parameters $\mathbf{g} \stackrel{\Delta}{=} (g_1, \ldots, g_J)$ and $\boldsymbol{\alpha} \stackrel{\Delta}{=} (\alpha_1, \ldots, \alpha_J)$, we shall use the notation $\Psi(\eta) \equiv \Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta)$.

Since the BER is a non-decreasing 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 $\eta^\star$ satisfies $\eta^\star g_1 \geq \text{SNR}(\epsilon)$, where the SNR level $\text{SNR}(\epsilon)$ is determined by the code $\mathcal{C}$. Let $0 \leq \delta_1 < \delta_2 \leq 1$ and $\delta_3 > 0$. We fix $g_1 = \text{SNR}(\epsilon)/\delta_2$ and obtain the other SNR levels $g_2, \ldots, g_J$ by sampling with a sufficiently small step a desired interval $[g_1, g_{\max}]$. Then, we look for the class load assignment $\boldsymbol{\alpha}$ solving the optimization program

$$\text{minimize} \quad \sum_{j=1}^J \alpha_j g_j \qquad \text{subject to} \left\{ \begin{array}{ll} \Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) \geq \eta + \delta_3, & \forall \ \ \eta \in [\delta_1, \delta_2] \\ \sum_{j=1}^J \alpha_j = \alpha, & \\ \alpha_j \geq 0, & \forall \ \ j \end{array} \right. \tag{36}$$

Suppose that (36) is feasible. Then, the solution $\boldsymbol{\alpha}^\star$ has the property of minimizing

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{\sum_{j=1}^J \alpha_j g_j}{2\alpha R}$$

over all class load assignments $\boldsymbol{\alpha}$ such that the spectral efficiency is equal to $\rho = \alpha R$, and the DE-GA has limit $\eta^\star \geq \delta_2$ (implying that all users attain BER not larger than ϵ). The parameter δ_3 governs the speed of convergence of DE-GA (and eventually of the true iterative decoder) to the fixed point. If δ_3 is very small, $\Psi(\mathbf{g}, \boldsymbol{\alpha}^\star, \eta)$ 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 observed in iterative decoding of Turbo Codes and LDPC codes through the so-called EXIT diagrams [38]). On the other hand, there is no hope to obtain small $(E_b/N_0)_{\text{sys}}$ by keeping δ_3 large. Therefore, δ_3 can be used as a performance vs. complexity tradeoff design parameter.

If for some α and ϵ the program (36) is infeasible, then some of the parameters must be changed, for example, by decreasing α and/or increasing the range $[g_1, g_{\max}]$ of permitted received SNR levels.

Fortunately, for all decoding algorithms considered in this paper, the condition $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) \geq \eta + \delta_3$ can be re-formulated as a linear constraint with respect to $\boldsymbol{\alpha}$. Therefore, (36) is a linear program and can be solved by standard numerical methods. Before proving the above statement, we would like to point out here that the optimization of $\boldsymbol{\alpha}$ via linear programming has striking analogy with the methods for optimizing the degree sequences of LDPC code ensembles, as for example in [35, 39].

For the SUMF-IC decoder, from (34) we re-write $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) \geq \eta + \delta_3$ as

$$\sum_{j=1}^{J} \alpha_j g_j \left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta))Dy\right) \leq \frac{1}{\eta + \delta_3} - 1 \tag{37}$$

which is clearly linear in $\boldsymbol{\alpha}$.

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

$$x \leq \Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) \;\Leftrightarrow\; x \leq f(\mathbf{g}, \boldsymbol{\alpha}, \eta, x) \tag{38}$$

Due to this iff implication, it follows that $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) \geq \eta + \delta_3$ is equivalent to

$$\sum_{j=1}^{J} \alpha_j g_j \int_{\mathbb{R}} \frac{\left(1 - \tanh^2(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta))\right)}{1 + (\eta + \delta_3)g_j\left(1 - \tanh^2(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta))\right)} Dy \leq \frac{1}{\eta + \delta_3} - 1 \tag{39}$$

for conditional MMSE-IC, and to

$$\sum_{j=1}^{J} \alpha_j g_j \frac{\left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta))Dy\right)}{1 + (\eta + \delta_3)g_j\left(1 - \int_{\mathbb{R}} \tanh^2(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta))Dy\right)} \leq \frac{1}{\eta + \delta_3} - 1 \tag{40}$$

for unconditional MMSE-IC. Again, both (39) and (40) are linear constraints in $\boldsymbol{\alpha}$. Finally, for the exact BP decoder we have to be a bit more careful because of the possibility of multiple solutions to the equation (29) defining the mapping function Ψ. Let us re-write (29) in the form $\Psi = f(\mathbf{g}, \boldsymbol{\alpha}, \eta, \Psi)$ and denote by $\Psi_1(\mathbf{g}, \boldsymbol{\alpha}, \eta) \leq \Psi_3(\mathbf{g}, \boldsymbol{\alpha}, \eta)$ its stable solutions. Clearly, the inequality $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) \geq \Psi_1(\mathbf{g}, \boldsymbol{\alpha}, \eta)$ always holds. Thus, the condition $\Psi_1(\mathbf{g}, \boldsymbol{\alpha}, \eta) \geq \eta + \delta_3$ for $\eta \in [\delta_1, \delta_2]$ implies the first constraint in (36).

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

$$f(\mathbf{g}, \boldsymbol{\alpha}, \eta,\, \psi) - \psi > 0, \quad \forall \psi \in [0,\, \eta + \delta_3),\; \forall \eta \in [\delta_1,\, \delta_2] \tag{41}$$

As desired, (41) is a collection of linear constraints on $\boldsymbol{\alpha}$, parametrized by ψ and η.

In practice, the linear constraints corresponding to (41) are obtained by sampling on an appropriate grid of points the trapezioidal region defined by $\psi \in [0, \eta + \delta_3)$, $\eta \in [\delta_1, \delta_2]$, in the (ψ, η)-plane. This may produce a large number of constraints. A simpler approach consists of requiring that (29) has a single solution. Then, if $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta) = \Psi_1(\mathbf{g}, \boldsymbol{\alpha}, \eta)$ is the unique solution of (29), then the condition (38) holds and the corresponding linear constraint is given by

$$\sum_{j=1}^{J} \alpha_j g_j \int_{\mathbb{R}^2} \frac{\left(1 - \tanh^2\left(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta)\right)\right)\left(1 - \tanh\left(z\sqrt{g_j(\eta+\delta_3)} + g_j(\eta+\delta_3)\right)\right)}{1 - \tanh^2\left(y\sqrt{\mu(g_j\eta)} + \mu(g_j\eta)\right)\tanh^2\left(z\sqrt{g_j(\eta+\delta_3)} + g_j(\eta+\delta_3)\right)} DzDy \leq \frac{1}{\eta+\delta_3} - 1 \tag{42}$$

By replacing the first constraint in (36) by (42), the vector $\boldsymbol{\alpha}^\star$ found by linear programming corresponds to a valid receiver SNR distribution if $\Psi = f(\mathbf{g}, \boldsymbol{\alpha}^\star, \eta, \Psi)$ 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 (42), finding a candidate $\boldsymbol{\alpha}^\star$ 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 $\boldsymbol{\alpha}^\star$ is unique.

As an example of the above optimization technique, consider Fig. 6, showing the DE-GA mapping function $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta)$ for the exact BP decoder, load $\alpha = 4.5$, maximum free-distance 64-state rate 1/3 convolutional user codes with generators $(133, 145, 175)_8$ (see [36]), and $(E_b/N_0)_{\rm sys} = 6$ dB. The curve corresponding to constant receiver SNR yields $\eta^\star \approx 0.1$, 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 $\eta^\star \approx 1.0$, i.e., each user attains its single-user performance after a sufficiently large number of iterations. The SNR-optimized curve in Fig. 6 is obtained by linear programming by using the constraint (42), enforced over grid of points in $[\delta_1 = 0, \delta_2 = 0.5]$, equally spaced by 0.01, and by letting $\delta_3 = 0.01$.

Fig. 7 shows the achievable spectral effciency 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 receiver SNR, with iterative detection and with separate detection (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 $(E_b/N_0)_{\rm sys} < 3.94$ 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 1, 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, 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 $(E_b/N_0)_{\rm sys}$, the iteratively-decoded systems with optimized SNR distribution are not interference limited, in the sense that their spectral efficiency increases with $(E_b/N_0)_{\rm sys}$. Remarkably, for the exact BP and the MMSE-IC decoders the large-$(E_b/N_0)_{\rm sys}$ slope of spectral efficiency is (close to) optimal, at least in the considered range of $(E_b/N_0)_{\rm sys}$.

- CDMA systems with constant receiver 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 $(E_b/N_0)_{\rm sys}$.

- The unconditional MMSE-IC yields spectral efficiency very close to exact BP with much smaller complexity with respect to both exact BP and conditional MMSE-IC. This makes the unconditional MMSE-IC decoder a good candidate for high-performance low-complexity iterative multiuser decoding. This point will be elaborated further in the next section.

5 Low-complexity implementation

In the previous section we showed that the unconditional MMSE-IC iterative decoder provides a good trade-off between spectral efficiency performance (under the optimized received SNR distribution) and complexity. Nevertheless, complexity is still fairly larger than conventional CDMA receivers, since it requires the computation of a bank of K MMSE filters (complexity $O(K^2)$ per user per iteration) at each decoder iteration. A solution reported in the literature [40] 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, given by

$$v_k = 1 - E\left[\tanh^2\left(\mathcal{L}_{k,n}^{\rm dec}/2\right)\right], \quad k = 1, \ldots, K \tag{43}$$

are below a certain threshold. This approach achieves complexity $O(K)$ per user per iteration, but it does not take into account the fact that, under optimal SNR distribution, users are received at different SNR levels and the evolution of their residual symbol variances with 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.

In order to illustrate the above intuition, consider the SNR distribution in Fig. 9, optimized by linear programming for the unconditional MMSE-IC receiver with $\alpha = 4.5$, convolutional code with generators $(133, 145, 175)_8$ and $(E_b/N_0)_{\rm sys} = 6.29$ dB. The distribution is composed by $J = 3$ SNR levels, denoted by g_1, g_2, g_3. Fig. 10 shows the evolution of the multiuser efficiency (left) and the residual user symbol variances (right) for the three classes of users vs. 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 was alone on the channel (the multiuser efficiency 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.

Fortunately, for the class of convolutional codes and iterative decoders considered in this paper and for a surprisingly large range of system parameters (user coding rates, $(E_b/N_0)_{\rm sys}$ and load α) the optimal SNR distribution consists of a small number J of discrete SNR levels, as in the example above. Next, we take advantage of this fact to obtain a low-complexity iterative multiuser decoding algorithm which performs very close to unconditional MMSE-IC with complexity $O(K)$, comparable to that of conventional CDMA receivers.[5]

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

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

[5]Our receiver algorithm applies 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). We 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, 41]).

where $\widehat{x}_{j,n}^{(\ell)}$ is the current estimate of the j-th user symbol given the SISO decoders output messages at the previous iteration, that for the binary antipodal case considered here is given by $\widehat{x}_{j,n} = \tanh(\mathcal{L}_{j,n}^{\rm dec}/2)$ (see (20)), and $\mathbf{f}_k^{(\ell)}$ is an appropriately chosen filter.[6]

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

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

where we define $\boldsymbol{\Gamma} \stackrel{\Delta}{=} \text{diag}(\gamma_1, \ldots, \gamma_K)$ and the residual symbol covariance matrix at iteration ℓ as

$$\mathbf{V}^{(\ell)} \stackrel{\Delta}{=} E\left[\left(\mathbf{x}_n - \widehat{\mathbf{x}}_n^{(\ell)}\right)\left(\mathbf{x}_n - \widehat{\mathbf{x}}_n^{(\ell)}\right)^T\right] = \text{diag}\left(v_1^{(\ell)}, \ldots, v_K^{(\ell)}\right) \tag{46}$$

Notice that $\mathbf{V}^{(\ell)}$ is *exactly* diagonal in the limit for large blocklength N because $\widehat{x}_{j,n}^{(\ell)}$ are obtained from the SISO decoders *extrinsic information* [17].

Under the optimized received SNR distribution, we shall assume that the users are grouped into $J \ll K$ classes of size $K_1, \ldots, K_J$. User k in class j is received at SNR level $\gamma_k = g_j$. As in Section 4, we let $g_1 \leq \cdots \leq g_J$ and enumerate the users such that users $k \in \{\mathcal{K}_{j-1}+1, \ldots, \mathcal{K}_j\}$ belongs to class j, where $\mathcal{K}_0 \stackrel{\Delta}{=} 0$ and $\mathcal{K}_j \stackrel{\Delta}{=} \sum_{i=1}^j K_i$.

The proposed low-complexity approach makes use of J linear detectors. Detector number j at iteration ℓ assumes user SNRs given by

$$u_{j,k}^{(\ell)} = \begin{cases} \xi^{(\ell)}\gamma_k v_k^{(\ell_j)} & \text{for } k = 1, 2, \ldots, \mathcal{K}_j \\ 0 & \text{for } k = \mathcal{K}_j + 1, \mathcal{K}_j + 2, \ldots, K \end{cases} \tag{47}$$

where $\xi^{(\ell)}$ is an iteration-dependent scaling factor common to all users (to be specified later) 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

$$\mathbf{U}_j^{(\ell)} = \xi^{(\ell)}\boldsymbol{\Gamma}\mathbf{V}^{(\ell_j)}\mathbf{Z}_j \tag{48}$$

Equation (48) is meaningful only for $\ell \geq \ell_j$. As it will be clear in the following, detectors are used in the order $j = J, J-1, \ldots, 1$ and the indices ℓ_j determine the detector switch points, i.e., the j-th detector is used for $\ell = \ell_j, \ldots, \ell_{j-1} - 1$, where $\ell_J = 0$ and ℓ_0 is the maximum number of iterations.

---

[6]We use $\mathbf{f}_k^{(\ell)}$ instead of $\mathbf{h}_k$ as in Section 2 in order to stress the fact that here the filter does not coincide necessarily 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.

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

$$\mathbf{S} = \mathbf{S}\mathbf{Z}_j + \underbrace{\mathbf{S}\left(\mathbf{I} - \mathbf{Z}_j\right)}_{\triangleq\, \widetilde{\mathbf{S}}_j} \tag{49}$$

and replace the *true* SNR diagonal matrix $\mathbf{\Gamma}\mathbf{V}^{(\ell)}$ in (45) by $\mathbf{U}_j^{(\ell)}$. We introduce the singular value decomposition

$$\mathbf{S}\left(\mathbf{\Gamma}\mathbf{V}^{(\ell_j)}\mathbf{Z}_j\right)^{1/2} = \mathbf{\Phi}_j\mathbf{D}_j\mathbf{\Theta}_j^T, \tag{50}$$

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

$$\mathbf{Q}_j \triangleq \mathbf{\Phi}_j^T\mathbf{S} \tag{51}$$

$$= \mathbf{D}_j\mathbf{\Theta}_j^T\mathbf{\Gamma}^{-1/2}\left(\mathbf{V}^{(\ell_j)}\right)^{-1/2} + \mathbf{\Phi}_j^T\widetilde{\mathbf{S}}_j. \tag{52}$$

Note that, though (52) looks more complicated, it may require fewer computing effort than (51) due to the diagonal structure of the matrices $\mathbf{D}_j, \mathbf{\Gamma}$, and $\mathbf{V}^{(\ell_j)}$ and the zero columns in $\widetilde{\mathbf{S}}_j$. By using (51) and (48) in (45) and in (44), we can write the j-th detector filter for user k at iteration $\ell = \ell_j, \ldots, \ell_{j-1} - 1$ as

$$\widetilde{\mathbf{f}}_{j,k}^{(\ell)} = \frac{\mathbf{\Phi}_j\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}\mathbf{q}_{j,k}}{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}\mathbf{q}_{j,k}} \tag{53}$$

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

$$\begin{aligned}
\widetilde{z}_{k,n}^{(\ell)} &= \frac{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}\mathbf{\Phi}_j^T}{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}\mathbf{q}_{j,k}}\left(\mathbf{y}_n - \mathbf{S}\mathbf{W}\widehat{\mathbf{x}}_n^{(\ell)} + \mathbf{s}_k w_k \widehat{x}_{k,n}^{(\ell)}\right) \\
&= \frac{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}}{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}\mathbf{q}_{j,k}}\left(\mathbf{\Phi}_j^T\mathbf{y}_n - \mathbf{Q}_j\mathbf{W}\widehat{\mathbf{x}}_n^{(\ell)} + \mathbf{q}_{j,k} w_k \widehat{x}_{k,n}^{(\ell)}\right) \\
&= \frac{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}}{\mathbf{q}_{j,k}^T\left[\xi^{(\ell)}\mathbf{D}_j^2 + \mathbf{I}\right]^{-1}\mathbf{q}_{j,k}}\underbrace{\left(\mathbf{\Phi}_j^T\mathbf{y}_n - \mathbf{Q}_j\mathbf{W}\widehat{\mathbf{x}}_n^{(\ell)}\right)}_{\triangleq \mathbf{d}_{n,j}^{(\ell)}} + w_k \widehat{x}_{k,n}^{(\ell)}
\end{aligned} \tag{54}$$

Given $\mathbf{Q}_j$ and the singular value decomposition (50), (54) has complexity $O(K)$ per user per iteration: notice that the calculation of $\mathbf{d}_{n,j}^{(\ell)}$ involves $O(K^2)/K$ operations per user, since it is common to all users. The other operations are just inner products of vectors with diagonal kernels. This brings the computational effort per user per iteration from quadratic to linear. Costly computations are needed only when a switch from detector j to detector $j+1$ takes place. Then, a singular value decomposition (50) and a matrix multiplication (51) or (52) are needed. The impact of this computations is not very large in typical situations with optimized SNR distribution. In the example at the end of this section, we have $J = 3$ swtich points and total iterations $\ell_0 = 55$, therefore the complexity of SVD per user per iteration yields $O(K^2)$, but it is multiplied by a factor $3/55$. In general, our approach is very effective for small J and a large number of decoder iterations (typical of heavily loaded systems attaining large spectral efficiency).

Two questions have been left open: how to determine the detector switch points ℓ_j and how to choose the scaling factor $\xi^{(\ell)}$. They will be addressed in the following.

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 $\widetilde{\mathbf{f}}_{j,k}^{(\ell)}$ 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 multiuser efficiency of an MMSE filter with power mismatch:

Proposition 7. Consider the CDMA system with K users and spreading factor L defined by $\mathbf{y} = \mathbf{S}\mathbf{P}^{1/2}\mathbf{x} + \boldsymbol{\nu}$ where $\boldsymbol{\nu} \sim \mathcal{N}(0, \mathbf{I})$ and the usual assumptions on random spreading made in Section 3 hold. Let $\mathbf{P} \triangleq \text{diag}(P_1, \ldots, P_K)$ such that $\max_k P_k \leq \overline{P}$, and let $(U_1, \ldots, U_K)$ be an arbitrary sequence of positive numbers such that $\min_k U_k \geq \underline{U}$, where $\overline{P}$ and $\underline{U}$ are fixed, finite and positive constants independent of K. Assume that, as $K \to \infty$, the joint empirical distribution of the pairs (P_k, U_k), defined by

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

converges almost everywhere to a given distribution $F_{P,U}(p, u)$. Then, by letting $K \to \infty$ with $K/L = \alpha$, the multiuser efficiency 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

$$\kappa = \eta \frac{1 + \alpha E_U\left[\frac{U}{(1+\eta U)^2}\right]}{1 + \alpha E_{P,U}\left[\frac{P}{(1+\eta U)^2}\right]} \tag{55}$$

where η is the solution to

$$\frac{1}{\eta} = 1 + \alpha E_U\left[\frac{U}{1 + \eta U}\right] \tag{56}$$

and where $E_U[\cdot]$ and $E_{P,U}[\cdot]$ 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 scketch of proof (skipping technicalities on almost-sure convergence) is given in Appendix B. □

Proposition 7 is the key for a simple and effective switching criterion. Before illustrating it, we take a short detour for the following interesting corollary. Notice that η given in (56) is the multiuser efficiency of a MMSE filter matched to powers $\{U_k\}$ rather than $\{P_k\}$. Hence, it can be regarded as the *nominal* multiuser efficiency of the mimatched detector, while κ given by (55) is the *true* multiuser efficiency 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 non-decreasing 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 7.1. 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* $\{U_k\}$.[7]

Proof. See Appendix B. □

Corollary 7.1 generalizes the result of [10] that, rephrased in our terminology, states that $\{U_k = \frac{1}{K}\sum_{j=1}^{K} P_j \,:\, k = 1,\dots,K\}$ 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].

Going back to the detector switching problem, 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 (55). For

[7] A 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 non-decreasing arrangements of $\{P_k\}$ and $\{U_k\}$, respectively.

detector j we have

$$\begin{aligned} \kappa_j^{(\ell)} &= \eta_j^{(\ell)} \frac{1 + \sum_{k=1}^{\mathcal{K}_j} \frac{\xi^{(\ell)}\gamma_k v_k^{(\ell_j)}}{(1+\eta_j^{(\ell)}\xi^{(\ell)}\gamma_k v_k^{(\ell_j)})^2}}{1 + \sum_{k=1}^{\mathcal{K}_j} \frac{\gamma_k v_k^{(\ell)}}{(1+\eta_j^{(\ell)}\xi^{(\ell)}\gamma_k v_k^{(\ell_j)})^2} + \sum_{k=\mathcal{K}_j+1}^{K} \gamma_k v_k^{(\ell)}} \\ \frac{1}{\eta_j^{(\ell)}} &= 1 + \sum_{k=1}^{\mathcal{K}_j} \frac{\xi^{(\ell)}\gamma_k v_k^{(\ell_j)}}{1+\eta_j^{(\ell)}\xi^{(\ell)}\gamma_k v_k^{(\ell_j)}} \end{aligned} \tag{57}$$

while for detector $j-1$ we have

$$\begin{aligned} \kappa_{j-1}^{(\ell)} &= \eta_{j-1}^{(\ell)} \frac{1 + \sum_{k=1}^{\mathcal{K}_{j-1}} \frac{\gamma_k v_k^{(\ell)}}{(1+\eta_{j-1}^{(\ell)}\gamma_k v_k^{(\ell)})^2}}{1 + \sum_{k=1}^{\mathcal{K}_{j-1}} \frac{\gamma_k v_k^{(\ell)}}{(1+\eta_{j-1}^{(\ell)}\gamma_k v_k^{(\ell)})^2} + \sum_{k=\mathcal{K}_{j-1}+1}^{K} \gamma_k v_k^{(\ell)}} \\ \frac{1}{\eta_{j-1}^{(\ell)}} &= 1 + \sum_{k=1}^{\mathcal{K}_{j-1}} \frac{\gamma_k v_k^{(\ell)}}{1+\eta_{j-1}^{(\ell)}\gamma_k v_k^{(\ell)}} \end{aligned} \tag{58}$$

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 (58) we implicitly assumed that $\xi^{(\ell)} = 1$ at each switch point (in particular, for $\ell = \ell_{j-1}$). Optimizing the detector with respect to $\xi^{(\ell)}$ appears to be a hard problem. A very effective heuristic choice enforcing the condition $\xi^{(\ell_j)} = 1$ for all $j = J, J-1, \ldots, 1$ is given by

$$\xi^{(\ell)} = \frac{\sum_{k=1}^{K} v_k^{(\ell)}}{\sum_{k=1}^{K} v_k^{(\ell_j)}}, \quad \text{for } \ell = \ell_j, \ell_j+1, \ldots, \ell_{j-1}-1 \tag{59}$$

In order to illustrate the behavior of the proposed low-complexity iterative scheme, we consider again the system and SNR distribution of Fig. 9. Fig. 11 shows the evolution of the multiuser efficiency (left) and the residual user symbol variances (right) for the three classes of users vs. the decoder iterations, for the low-complexity algorithm described above. 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 of 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.

6 Concluding remarks

In this work we have extended the DE-GA analysis approach of iterative multiuser joint decoding of [17] to the exact 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 non-uniform symbols prior probabilities. Based on the DE-GA performance characterization of the exact BP iterative decoder and of its IC-based approximations, we have formulated the problem of optimal user received SNR distribution in terms of simple linear programming. This allowed us to compute 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 $(E_b/N_0)_{\rm sys}$. Interestingly, the simple unconditional MMSE-IC algorithm performs very close to the exact 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 (essentially linear in the number of users or, equivalently, *constant* per decoded information bit) with very small degradation. As a byproduct of the development of the proposed low-complexity algorithm, we obtained an interesting general expression for the multiuser efficiency of an MMSE filter with mismatched user powers, in the large-system limit.

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 $(E_b/N_0)_{\rm sys}$, 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 multi-cell environment the optimal (single-cell) SNR distribution is expected to provide a smaller total emitted energy from each cell, thus reducing the inter-cell 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 inter-cell 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 develped in [9] for uniform symbol a priori probabilities and constant user power, and extended in [43] 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 [44]. 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 $\{\mathcal{L}_{j,n}^{\rm dec} : j \neq k\}$ 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 $K \to \infty$, the empirical distribution of these prior probabilities converges almost everywhere to some deterministic distirbution. 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), the pdf of the channel output $\mathbf{y}_n$ conditioned on the spreading sequences $\mathbf{S}$ is proportional to

$$Z(\mathbf{y}_n, \mathbf{S}) = \sum_{\mathbf{x}_n \in \{\pm 1\}^K} \Pr(\mathbf{x}_n) \exp\left(-\frac{1}{2\sigma^2}|\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{x}_n|^2\right) \tag{60}$$

if the fictious noise variance σ^2 is set to the true noise variance σ_0^2. Moreover (60) is independent of n since the input is assumed to be 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

$$\mathcal{F}_K(\mathbf{y}, \mathbf{S}) = \frac{1}{K} \log Z(\mathbf{y}, \mathbf{S}) \tag{61}$$

is called the *free energy*. One of the fundamental principles of statistical mechanics is that the free energy is self-averaging in the large system limit. That is,

$$\lim_{K\to\infty} \mathcal{F}_K(\mathbf{y}, \mathbf{S}) = \lim_{K\to\infty} E[\mathcal{F}_K(\mathbf{y}, \mathbf{S})] \stackrel{\Delta}{=} \mathcal{F} \tag{62}$$

with probability 1, where averaging is with respect to the random spreading sequences and the channel noise. A standard trick used in statistical mechanics in order to compute $\mathcal{F}$ is the Replica Method. This consists of re-writing the free energy in the following way

$$\mathcal{F} = \lim_{K\to\infty} \frac{1}{K} \lim_{n\to 0} \frac{\partial}{\partial n} \log\left(E[Z^n(\mathbf{y}, \mathbf{S})]\right) \tag{63}$$

with the advantage that the expectation operator has moved into the argument of the logarithm. Now, the free energy is evaluated for integer n and the results is assumed to generalize to positive real n. Further discussion about the Replica Method and its justification is provided in [44, 9].

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 nat) per dimension is given by

$$\mathcal{I} = -\alpha\, \mathcal{F}|_{\sigma^2=\sigma_0^2} + \frac{1}{2}\log 2\pi\sigma_0^2 - \frac{1}{2}\log 2\pi e\sigma_0^2 = -\alpha\, \mathcal{F}|_{\sigma^2=\sigma_0^2} - \frac{1}{2} \tag{64}$$

Following [9], by using the *Replica Symmetry* assumption (see [9, 43]), Cramér Theorem and Varadhan Lemma [45], we get that the free energy can be expressed as follows. Let $\xi(n)$ be defined by the following saddle-point

$$\xi(n) = \sup_{m,q} \inf_{\widetilde{m},\widetilde{q}} \left\{ \frac{1}{\alpha} G(m, q) - n\widetilde{m}m - \frac{n(n-1)}{2}\widetilde{q}q + E_{V,\mathcal{L}}\left[\log \Phi(V\widetilde{m}, V\widetilde{q}, \mathcal{L})\right] \right\} \tag{65}$$

where n is the replica order. Then,

$$\mathcal{F} = \lim_{n\to 0} \frac{d}{dn}\xi(n) \tag{66}$$

In (65), $V, \mathcal{L}$ are random variables with joint distribution equal to the limit of the empirical joint distribution of

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

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

$$\{x_a x_b : a = 1, \ldots, n, \quad 0 \le b < a\}$$

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

$$\lambda_{ab} = \begin{cases} \widetilde{m} & b = 0, a \neq b \\ \widetilde{q} & b > 0, a \neq b \end{cases}$$

This is given by

$$\begin{aligned}
\Phi(\widetilde{m}, \widetilde{q}, \mathcal{L}) &= \sum_{(x_0,x_1,\ldots,x_n)\in\{\pm 1\}^{n+1}} \left(\prod_{i=0}^{n} \Pr(x_i)\right) \exp\left(\sum_{b<a} x_a x_b \lambda_{ab}\right) \\
&= \sum_{(x_0,x_1,\ldots,x_n)\in\{\pm 1\}^{n+1}} \left(\prod_{i=0}^{n} \Pr(x_i)\right) \exp\left(\widetilde{m} x_0 \sum_{a=1}^{n} x_a + \frac{\widetilde{q}}{2}\left(\sum_{a=1}^{n} x_a\right)^2 - \frac{n\widetilde{q}}{2}\right) \\
&= \sum_{(x_1,\ldots,x_n)\in\{\pm 1\}^{n}} \left(\prod_{i=1}^{n} \Pr(x_i)\right) \left(\frac{1+T}{2} \exp\left(\widetilde{m} \sum_{a=1}^{n} x_a + \frac{\widetilde{q}}{2}\left(\sum_{a=1}^{n} x_a\right)^2\right) + \right. \\
&\quad \left. + \frac{1-T}{2} \exp\left(-\widetilde{m} \sum_{a=1}^{n} x_a + \frac{\widetilde{q}}{2}\left(\sum_{a=1}^{n} x_a\right)^2\right)\right) e^{-n\widetilde{q}/2} \qquad (67) \\
&= \sum_{(x_1,\ldots,x_n)\in\{\pm 1\}^{n}} \left(\prod_{i=1}^{n} \Pr(x_i)\right) \left(\frac{1+T}{2} \int_{\mathbb{R}} \exp\left((z\sqrt{\widetilde{q}} + \widetilde{m}) \sum_{a=1}^{n} x_a\right) Dz + \right. \\
&\quad \left. + \frac{1-T}{2} \int_{\mathbb{R}} \exp\left(-(z\sqrt{\widetilde{q}} + \widetilde{m}) \sum_{a=1}^{n} x_a\right) Dz\right) e^{-n\widetilde{q}/2} \qquad (68)
\end{aligned}$$

where we have used the identity (Hubbard-Stratonovich transform [9])

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

and we have defined $T = \tanh(\mathcal{L}/2)$ 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 Gaussian Approximation, we have that for 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, for later use we notice that, from (67), we have

$$\frac{\partial}{\partial \widetilde{q}} \Phi(\widetilde{m}, \widetilde{q}, \mathcal{L}) = \frac{1}{2} \frac{\partial^2}{\partial \widetilde{m}^2} \Phi(\widetilde{m}, \widetilde{q}, \mathcal{L}) - \frac{n}{2} \Phi(\widetilde{m}, \widetilde{q}, \mathcal{L}) \qquad (69)$$

We can further simplify the expression of the moment-generating function in (68) by noticing that

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

and, similarly,

$$\Pr(x_a = +1)e^{-(z\sqrt{\widetilde{q}}+\widetilde{m})} + \Pr(x_a = -1)e^{z\sqrt{\widetilde{q}}+\widetilde{m}} = \frac{\cosh(z\sqrt{\widetilde{q}}+\widetilde{m}-\mathcal{L}/2)}{\cosh(\mathcal{L}/2)}$$

Then, by rearranging the terms in the sum with respect to $x_1, \ldots, x_n$ in (68) and summing by using the two above identities, we obtain

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

The function $G(m,q)$ in (65) does not depend on the symbols prior probabilities and is directly obtained from [9] as

$$G(m,q) = \frac{1}{2}\log\frac{(1+\frac{\alpha}{\sigma^2}(1-q))^{n-1}}{1+\frac{\alpha}{\sigma^2}(1-q)+\frac{n\sigma_0^2}{\sigma^2}(1+\frac{\alpha}{\sigma_0^2}(1-2m+q))} \tag{71}$$

Let $F(n,m,q,\widetilde{m},\widetilde{q})$ denote the argument of the extremization (sup-inf) in (65). The saddle-point condition is obtained by the set of equations

$$\frac{\partial}{\partial m}F = 0, \quad \frac{\partial}{\partial q}F = 0, \quad \frac{\partial}{\partial \widetilde{m}}F = 0, \quad \frac{\partial}{\partial \widetilde{q}}F = 0 \tag{72}$$

In order to obtain $\mathcal{F}$ we should: 1) find the solution of the system of equations (72); 2) find $\xi(n)$ by substituting this solution into $F(n,m,q,\widetilde{m},\widetilde{q})$; 3) evaluate the derivative and the limit in (66). As a matter of fact, from a continuity argument, it is equivalent (but easier) to solve for the system of equations

$$\lim_{n\to 0}\frac{\partial}{\partial m}F = 0, \quad \lim_{n\to 0}\frac{\partial}{\partial q}F = 0, \quad \lim_{n\to 0}\frac{\partial}{\partial \widetilde{m}}F = 0, \quad \lim_{n\to 0}\frac{\partial}{\partial \widetilde{q}}F = 0 \tag{73}$$

and substitute the solution into

$$\lim_{n\to 0}\frac{\partial}{\partial n}F(n,m,q',\widetilde{m},\widetilde{q})$$

After some algebra, using (69), the saddle-point equations (73) can be put in the explicit form

$$\begin{aligned} m &= E_{V,\mathcal{L}}\left[V\int_{\mathbb{R}}\left\{\frac{1+T}{2}\tanh\left(z\sqrt{V\widetilde{q}}+V\widetilde{m}+\mathcal{L}/2\right)+\frac{1-T}{2}\tanh\left(z\sqrt{V\widetilde{q}}+V\widetilde{m}-\mathcal{L}/2\right)\right\}Dz\right] \\ q &= E_{V,\mathcal{L}}\left[V\int_{\mathbb{R}}\left\{\frac{1+T}{2}\tanh^2\left(z\sqrt{V\widetilde{q}}+V\widetilde{m}+\mathcal{L}/2\right)+\frac{1-T}{2}\tanh^2\left(z\sqrt{V\widetilde{q}}+V\widetilde{m}-\mathcal{L}/2\right)\right\}Dz\right] \\ \widetilde{m} &= \frac{1}{\sigma^2+\alpha(1-m)}, \qquad \widetilde{q}=\frac{\sigma_0^2+\alpha(1-2m+q)}{(\sigma^2+\alpha(1-m))^2} \end{aligned} \tag{74}$$

Since we are interested in the IO-MUD, we let $\sigma^2 = \sigma_0^2$ (see [9]) and find that, at the saddle-point given by (74), the solution yields $m = q$ and $\widetilde{m} = \widetilde{q}$. In this case, by following the gauge-theory argument of [46], it is possible to show that the Replica Symmetry assumption made before in order to obtain (65) is indeed valid for the stable solutions of the IO-MUD with arbitrary symbol priors.

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: 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. In a way completely analogous to [9], we can show that in the large-system limit

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

Hence, $\sigma_0^2\widetilde{m}$ (independent of the user index k) is the ME of the IO-MUD in the large-system limit. By using the first and the third equation in the system (74), and replacing q and $\widetilde{q}$ by m and $\widetilde{m}$, respectively, and by substituting $\widetilde{m} = \Psi/\sigma_0^2$ we obtain the implicit expression for the ME of the IO-MUD with arbitrary prior probabilities on the symbols of the interfering users as

$$\begin{aligned} \frac{1}{\Psi} &= 1+\alpha\left(\frac{1}{\sigma_0^2}-E_{\gamma,\mathcal{L}}\left[\gamma\int_{\mathbb{R}}\left\{\frac{1+T}{2}\tanh\left(z\sqrt{\gamma\Psi}+\gamma\Psi+\frac{\mathcal{L}}{2}\right)+\right.\right.\right. \\ &\qquad \left.\left.\left.\frac{1-T}{2}\tanh\left(z\sqrt{\gamma\Psi}+\gamma\Psi-\frac{\mathcal{L}}{2}\right)\right\}Dz\right]\right) \\ &= 1+\alpha E_{\gamma,\mathcal{L}}\left[\gamma\left(1-T^2\right)\int_{\mathbb{R}}\frac{1-\tanh\left(z\sqrt{\gamma\Psi}+\gamma\Psi\right)}{1-T^2\tanh^2\left(z\sqrt{\gamma\Psi}+\gamma\Psi\right)}Dz\right] \end{aligned} \tag{75}$$

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 (75) we obtain (29). By computing $\lim_{n\to 0}\frac{\partial}{\partial n}F(n, m, q', \widetilde{m}, \widetilde{q})$ in the saddle-point solution given by (74), 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 bit/dimension.

From standard results in statistical mechanics [9], when (74) has multiple stable solutions, the performance of the IO-MUD corresponds to the solution maximizing the free energy or, equivalently, minimizing the mutual information. This proves all statements in Proposition 2.

B Proofs of Proposition 7 and of Corollary 7.1

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

$$\begin{aligned} \boldsymbol{\Sigma}_U &= \mathbf{I} + \sum_{j\neq k} U_j \mathbf{s}_j \mathbf{s}_j^T \\ \boldsymbol{\Sigma}_P &= \mathbf{I} + \sum_{j\neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \end{aligned} \tag{76}$$

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

$$\beta_k = P_k \frac{(\mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k)^2}{\mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \boldsymbol{\Sigma}_P \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k} \tag{77}$$

From the standard result of [6] we have that, under the assumptions of Proposition 7, 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 (56).

We re-write the denominator of (77) in the large-system limit as

$$\begin{aligned} \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \boldsymbol{\Sigma}_P \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k &= \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-1} \left(\sum_{j\neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \right) \boldsymbol{\Sigma}_U^{-1} \mathbf{s}_k + \mathbf{s}_k^T \boldsymbol{\Sigma}_U^{-2} \mathbf{s}_k \\ &\stackrel{\text{w.p. 1}}{\to} \lim_{K\to\infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_U^{-1} \left(\sum_{j\neq k} P_j \mathbf{s}_j \mathbf{s}_j^T \right) \boldsymbol{\Sigma}_U^{-1} \right) + \lim_{K\to\infty} \frac{1}{L} \text{tr} \left(\boldsymbol{\Sigma}_U^{-2} \right) \end{aligned} \tag{78}$$

where we used the result proved in [47]

$$\lim_{K\to\infty} \mathbf{s}_k^T \mathbf{A}_K \mathbf{s}_k \stackrel{\text{w.p. 1}}{=} \lim_{K\to\infty} \frac{1}{L} \text{tr}(\mathbf{A}_K) \tag{79}$$

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 (78), the maximum eigenvalue is always upper bounded by 1.

Next, we evaluate the two limits in (78) separately. For the first we have

$$\begin{aligned}
\frac{1}{L}\mathrm{tr}\left(\boldsymbol{\Sigma}_U^{-1}\left(\sum_{j\neq k} P_j \mathbf{s}_j\mathbf{s}_j^T\right)\boldsymbol{\Sigma}_U^{-1}\right) &= \frac{1}{L}\sum_{j\neq k} P_j \mathbf{s}_j^T\boldsymbol{\Sigma}_U^{-2}\mathbf{s}_j \\
&\stackrel{\mathrm{a}}{=} \frac{1}{L}\sum_{j\neq k} P_j \mathbf{s}_j^T\left[\boldsymbol{\Sigma}_{U,j}^{-1} - \frac{U_j}{1+U_j\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-1}\mathbf{s}_j}\boldsymbol{\Sigma}_{U,j}^{-1}\mathbf{s}_j\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-1}\right]^2\mathbf{s}_j \\
&= \frac{1}{L}\sum_{j\neq k} P_j\left[\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-2}\mathbf{s}_j + \left(\frac{U_j\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-1}\mathbf{s}_j}{1+U_j\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-1}\mathbf{s}_j}\right)^2\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-2}\mathbf{s}_j - \right. \\
&\quad \left. - \left(\frac{2U_j\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-1}\mathbf{s}_j}{1+U_j\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-1}\mathbf{s}_j}\right)\mathbf{s}_j^T\boldsymbol{\Sigma}_{U,j}^{-2}\mathbf{s}_j\right] \\
&\stackrel{\mathrm{w.p.}\ 1}{\to} \lim_{K\to\infty}\frac{\alpha}{K}\sum_{j\neq k}\frac{P_j B}{(1+U_jA)^2} \\
&= \alpha E_{U,P}\left[\frac{PB}{(1+UA)^2}\right] \qquad (80)
\end{aligned}$$

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 repeadetly the lemma (79), and we defined the limits

$$\begin{aligned}
A &= \lim_{K\to\infty}\frac{1}{L}\mathrm{tr}\left(\boldsymbol{\Sigma}_{U,j}^{-1}\right) \\
B &= \lim_{K\to\infty}\frac{1}{L}\mathrm{tr}\left(\boldsymbol{\Sigma}_{U,j}^{-2}\right) \qquad (81)
\end{aligned}$$

Under the assumptions of Proposition 7, the matrix $\boldsymbol{\Sigma}_U, \boldsymbol{\Sigma}_{U,j}$ and $\boldsymbol{\Sigma}_P$ have all a well-defined limiting eigenvalue distribution and are invertible with probability 1, therefore, the limits in (81) exist and are immediately obtained from the limiting eigenvalue distributions.

For the second limit in (78) we have, again using lemma (79),

$$\lim_{K\to\infty}\frac{1}{L}\mathrm{tr}\left(\boldsymbol{\Sigma}_U^{-2}\right) = B \qquad (82)$$

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

$$\begin{aligned}
A &= \int \frac{1}{1+\lambda}dG(\lambda) \\
B &= \int \frac{1}{(1+\lambda)^2}dG(\lambda) \qquad (83)
\end{aligned}$$

Eventually, the last line of (78) is given by

$$D = \alpha E_{U,P}\left[\frac{P\int\frac{1}{(1+\lambda)^2}dG(\lambda)}{\left(1+U\int\frac{1}{1+\lambda}dG(\lambda)\right)^2}\right] + \int\frac{1}{(1+\lambda)^2}dG(\lambda) \tag{84}$$

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

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

We have

$$m_G(z) = \frac{1}{-z + \alpha E_U\left[\frac{U}{1+Um_G(z)}\right]} \tag{85}$$

In particular, we have that η defined in (56) is given by

$$\eta = \int\frac{1}{1+\lambda}dG(\lambda) = m_G(-1) \tag{86}$$

Furthermore,

$$\int\frac{1}{(1+\lambda)^2}dG(\lambda) = \left.\frac{d}{dz}m_G(z)\right|_{z=-1} \triangleq m'_G(-1) \tag{87}$$

By substituting (86) and (87) in the SINR denominator (84), and using the result in the expression for the SINR (77) we can write the large-system ME of user k, given by β_k/P_k, as

$$\kappa = \frac{\eta^2}{m'_G(-1)\left(1+\alpha E_{U,P}\left[\frac{P}{(1+\eta U)^2}\right]\right)} \tag{88}$$

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

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

By using the above expression in (88) we obtain (55). This concludes the proof of Proposition 7.

Next, we focus on the proof of Corollary 7.1. 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 non-decreasing order by the same permutation. Without loss of generality, we assume that $\{P_k\}$ and $\{U_k\}$ are non-decreasing (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 7, respectively. From (55) we see that this is verified if and only if

$$E_U\left[\frac{U}{(1+\eta U)^2}\right] \geq E_{P,U}\left[\frac{P}{(1+\eta U)^2}\right] \tag{89}$$

Assume, for the time being, a finite number of users. Then, inequality (89) becomes

$$\frac{1}{K}\sum_{k=1}^{K}\frac{U_k}{(1+\eta U_k)^2} \geq \frac{1}{K}\sum_{k=1}^{K}\frac{P_k}{(1+\eta U_k)^2} \tag{90}$$

We make use of the following lemma, proved in [48]:

Lemma. Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be real vectors of dimension K with non-decreasing components. If $\mathbf{b}$ majorizes $\mathbf{a}$, then

$$\sum_{k=1}^{K}\frac{a_k}{c_k} \geq \sum_{k=1}^{K}\frac{b_k}{c_k} \tag{91}$$

□

We apply the lemma to (90) 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 (90) is verified. Finally, a standard continuity argument extends the result to $K \to \infty$.

References

[1] S. Verdú, *Multiuser detection*, Cambridge University Press, Cambridge, UK, 1998.

[2] T. Cover and J. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[3] R. Gallager, *Information theory and reliable communication*, Wiley, New York, 1968.

[4] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 622–640, March 1999.

[5] S. Verdú and S. Shamai, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Inform. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[6] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and capacity," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 641–675, March 1999.

[7] D. Tse and S. Verdú, "Optimum asymptotic multiuser efficiency of randomly spread CDMA," *IEEE Trans. on Inform. Theory*, vol. 46, no. 11, pp. 2718–2722, November 2000.

[8] E. Chong, J. Zhang, and D. Tse, "Output MAI distribution of linear MMSE multiuser receivers in DS-CDMA systems," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 1128–1144, March 2001.

[9] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. on Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, November 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, no. 8, pp. 1429–1441, August 2001.

[11] G. Caire and R. Müller, "The optimal received power distribution for ic-based iterative multiuser joint decoders," in *Proc. 39th Allerton Conf. Commun. Contr. and Comp.*, Monticello, IL, USA, Oct. 2001.

[12] R. Müller and G. Caire, "Efficient implementation of iterative multiuser decoding," in *Proc. ISIT 2002*, Lausanne, Switzerland, July 2002.

[13] G. Caire, R. Müller, and T. Tanaka, "Density evolution and power profile optimization for iterative multiuser decoders based on individually optimum multiuser detectors," in *Proc. 40th Allerton Conf. Commun. Contr. and Comp.*, Monticello, IL, USA, Oct. 2002.

[14] Kiran and D. Tse, "Effective bandwidths and effective interference for linear multiuser receivers in asynchronous systems," *IEEE Trans. on Inform. Theory*, vol. 46, no. 4, pp. 1426–1447, July 2000.

[15] J. Evans and D. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. on Inform. Theory*, vol. 46, no. 6, pp. 2059–2078, September 2000.

[16] A. J. Viterbi, *CDMA – Principles of spread spectrum communications*, Addison-Wesley, Reading, MA, 1995.

[17] J. Boutros and G. Caire, "Iterative multiuser decoding: unified framework and asymptotic performance analysis," *IEEE Trans. on Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, July 2002.

[18] 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 Conference*, Pacifi c Groove, CA, November 1997.

[19] A. Roumy, S. Guemghar, G. Caire, and S. Verdú, "Maximizing the spectral effi ciency of LDPC-encoded CDMA," submitted to IEEE Trans. on Inform. Theory, September 2002.

[20] R. M¨uller and W. Gerstacker, "On the capacity loss due to separatation of detection and decoding," submitted to IEEE Trans. on Inform. Theory, 2002.

[21] "Special issue on iterative decoding," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, February 2001.

[22] X. Wang and V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. on Commun.*, vol. 47, no. 7, pp. 1047–1061, July 1999.

[23] P. Alexander, A. Grant, and M. Reed, "Iterative detection in code-division multiple-access with error control coding," *European Trans. on Telecomm.*, vol. 9, no. 5, pp. 419–425, September 1999.

[24] H. ElGamal and E. Geraniotis, "Iterative multiuser detection for coded cdma signals in awgn and fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 1, pp. 30–41, January 2000.

[25] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message passing decoding," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[26] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, San Francisco, 1988.

[27] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 498–519, February 2001.

[28] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Inform. Theory*, vol. 20, no. 3, pp. 284–287, March 1974.

[29] E. Viterbo and J. Boutros, "A universal lattice code decoder fro fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 6, pp. 1639–1542, July 1999.

[30] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple antenna channel," submitted to IEEE Trans. on Inform. Theory, 2001.

[31] V. Poor, *An introduction to signal detection and estimation*, Springer-Verlag, New York, 1988.

[32] D. Divsalar, M. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. on Commun.*, vol. 46, no. 2, pp. 258–268, February 1998.

[33] T. Richardson, R. Urbanke, and A. Shokrollahi, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[34] H. El Gamal and R. Hammons, "Analyzing the turbo decoder using the Gaussian approximation," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 671–686, February 2001.

[35] Sae Young Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 657–670, February 2001.

[36] J. Proakis, *Digital communications, 3rd Ed.*, McGraw-Hill, New York, 1995.

[37] Y. Kuznetsov, *Elements of applied bifurcation theory*, Springer-Verlag, New York, 1998.

[38] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Commun.*, vol. 49, no. 10, pp. 1727–1737, October 2001.

[39] A. Roumy, S. Guemghar, G. Caire, and S. Verdú, "Design methods for irregular repeat-accumulate codes," submitted to IEEE Trans. on Inform. Theory, October 2002.

[40] 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. on Commun. (ICC)*, Helsinki, Finland, June 2001.

[41] L. Li, A. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Communications in Information and Systems*, vol. 1, pp. 273–304, Fall 2001.

[42] A. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications*, Academic Press, San Diego, 1979.

[43] D. Guo and S. Verdú, "Multiuser detection and statistical mechanics," in *Ian Blake Festschrift*, 2002.

[44] H. Nishimori, *Statistical physics of spin glasses and information processing*, Oxford University Press, Oxford, U.K., 2001.

[45] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer, 2nd ed., New York, 1998.

[46] H. Nishimori, "Comment on "Statistical mechanics of CDMA multiuser demodulation" by T. Tanaka," *Europhys. Lett.*, vol. 57, no. 2, pp. 302–303, 2002.

[47] V. Marcenko and L. Pastur, "Distribution of eigenvalues for some sets of random matrices," Math. USSR, Sb., , no. 1, pp. 457–483, 1967.

[48] S. Ali Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," to appear on IEEE Trans. on Wireless Commun., 2003.

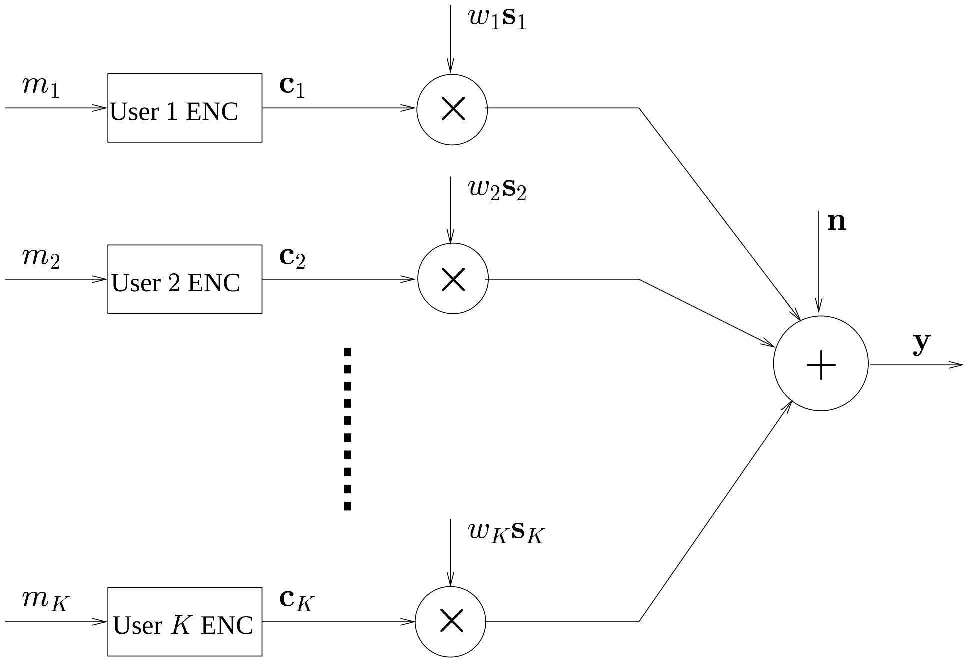

Figure 1: Coded CDMA with AWGN. User encoders may include interleaving.

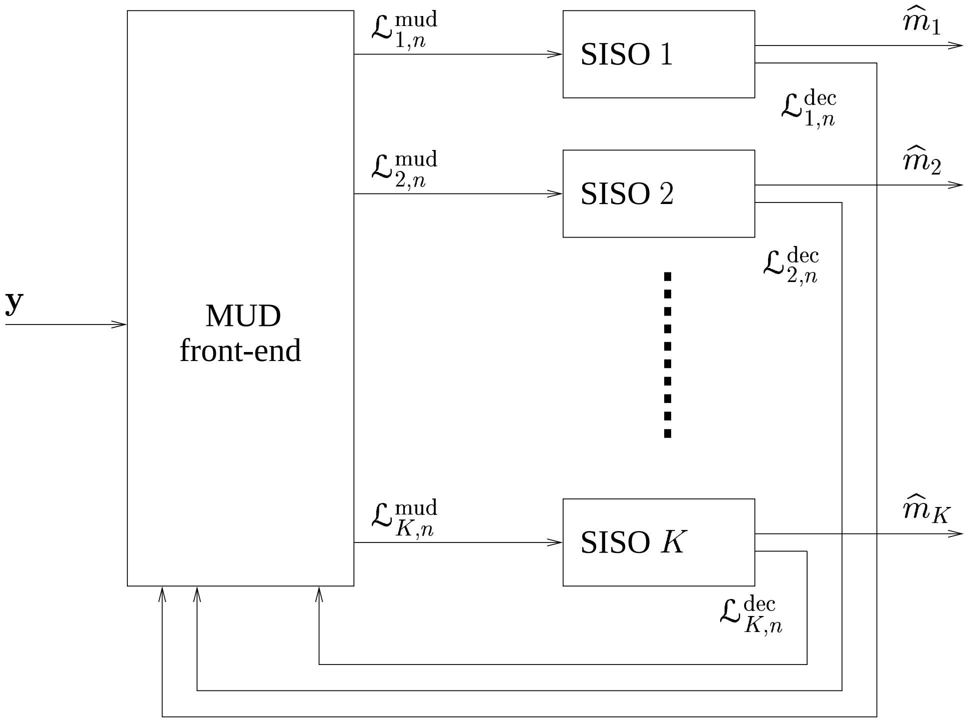

Figure 2: Block diagram of a multiuser iterative joint decoder.

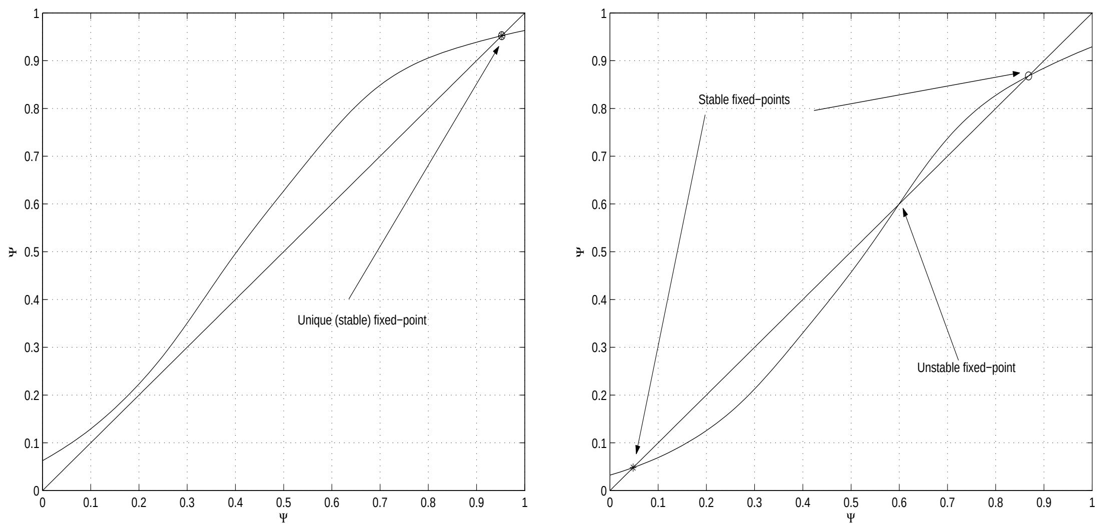

Figure 3: The curves show the reciprocal of the RHS of (29) as a function of $\Psi \in [0, 1]$ for $\alpha = 1.5$ (left), $\alpha = 3.0$ (right) constant $\gamma = 10$ dB and $\eta = 0$. The intersections of the curves with the straight line give the solutions of (29).

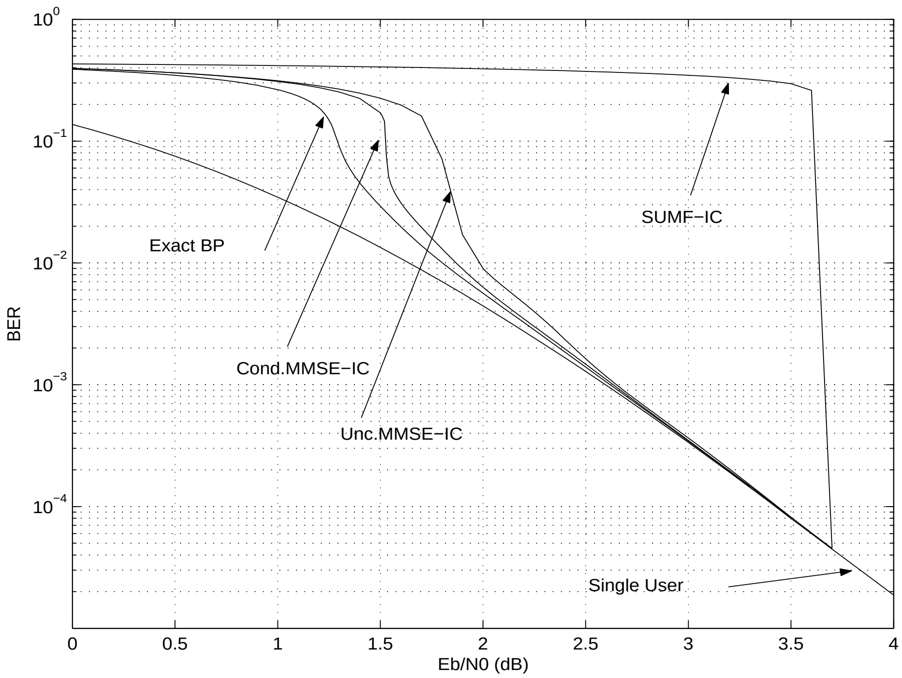

Figure 4: BER vs. E_b/N_0 for $\alpha = 1.0$, constant receiver SNR, convolutional code with generators $(133, 171)_8$ and different iterative decoding algorithms.

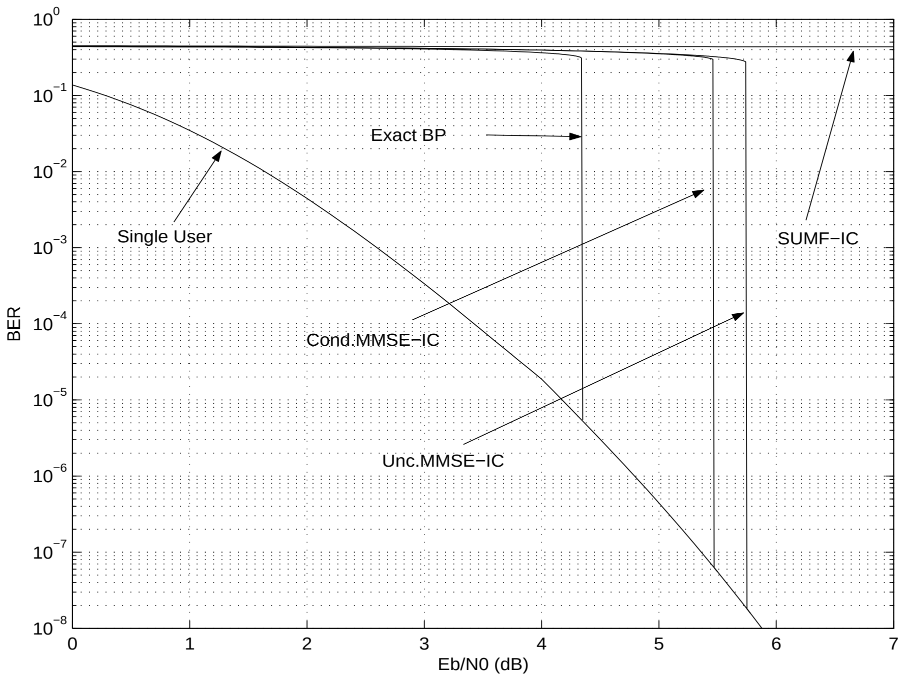

Figure 5: BER vs. E_b/N_0 for $\alpha = 2.0$, constant receiver SNR, convolutional code with generators $(133, 171)_8$ and different iterative decoding algorithms.

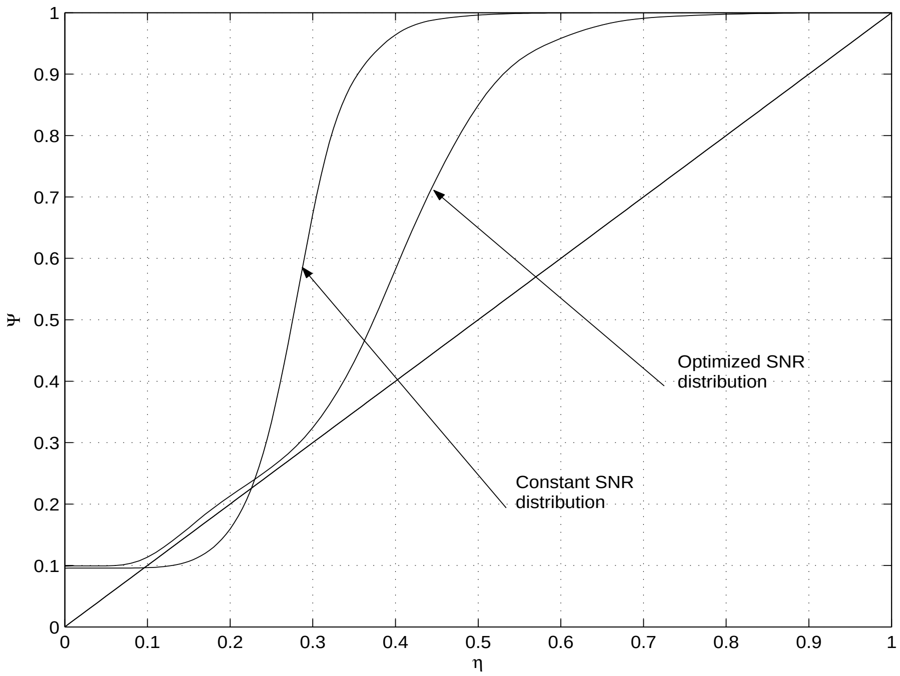

Figure 6: DE-GA mapping function $\Psi(\mathbf{g}, \boldsymbol{\alpha}, \eta)$ for the exact BP decoder, $E_b/N_0 = 6$ dB, constant and optimized SNR distributions, $\alpha = 4.5$ and convolutional code with generators $(133, 145, 175)_8$.

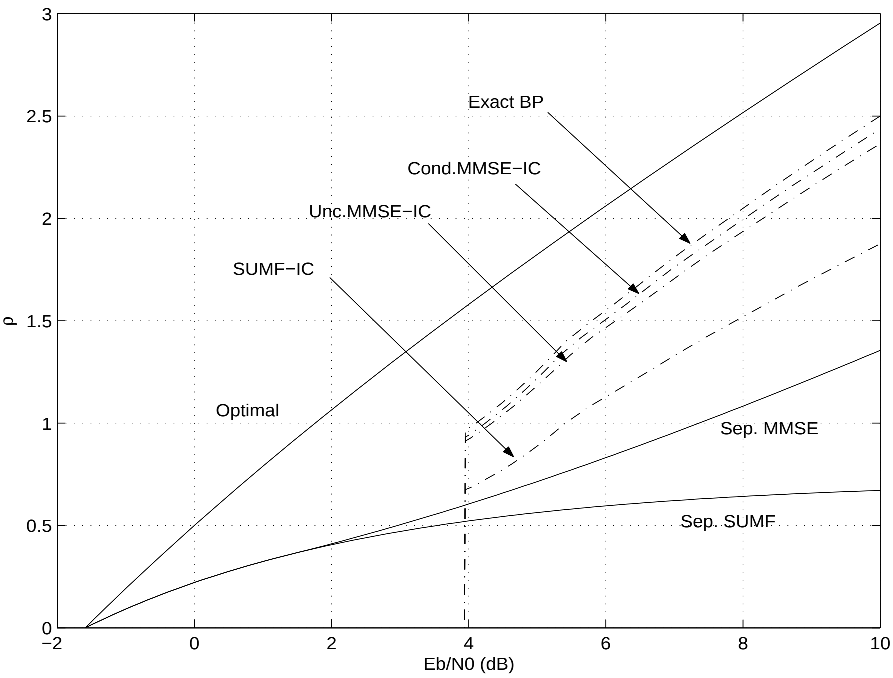

Figure 7: Spectral efficiency vs. E_b/N_0 at BER $\leq 10^{-5}$ for convolutionally-code CDMA with user codes with generators $(133, 145, 175)_8$, optimized received SNR distribution and different iterative decoding algorithms. Curves for joint detection and optimal codes (binary and Gaussian input),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

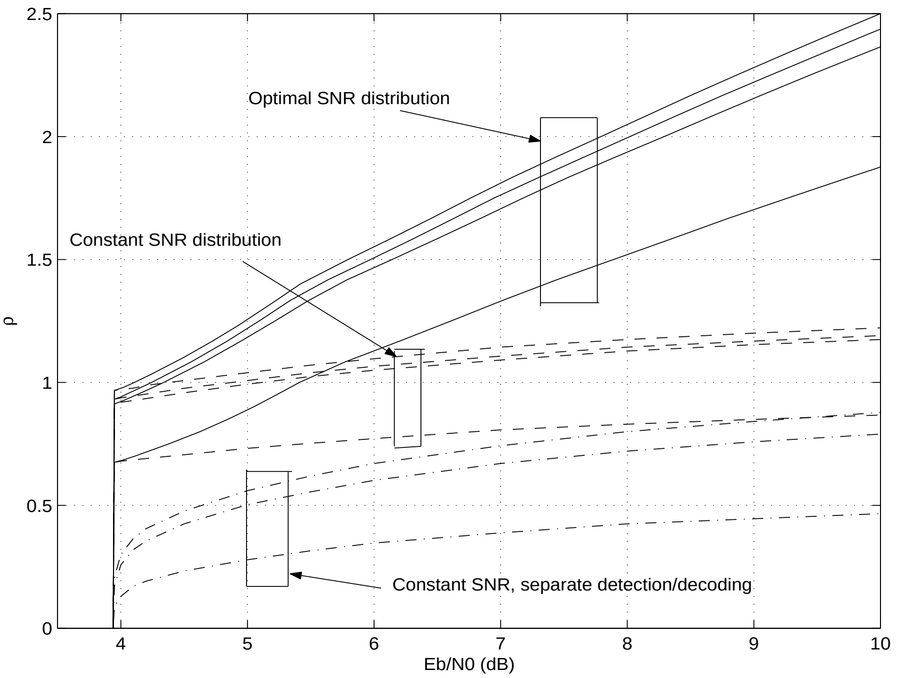

Figure 8: Spectral efficiency vs. E_b/N_0 at BER $\leq 10^{-5}$ for convolutionally-code CDMA with user codes with generators $(133, 145, 175)_8$, optimized received SNR distribution and constant received SNR distribution with iterative and separate detectiona and decoding. Each set of curves shows the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

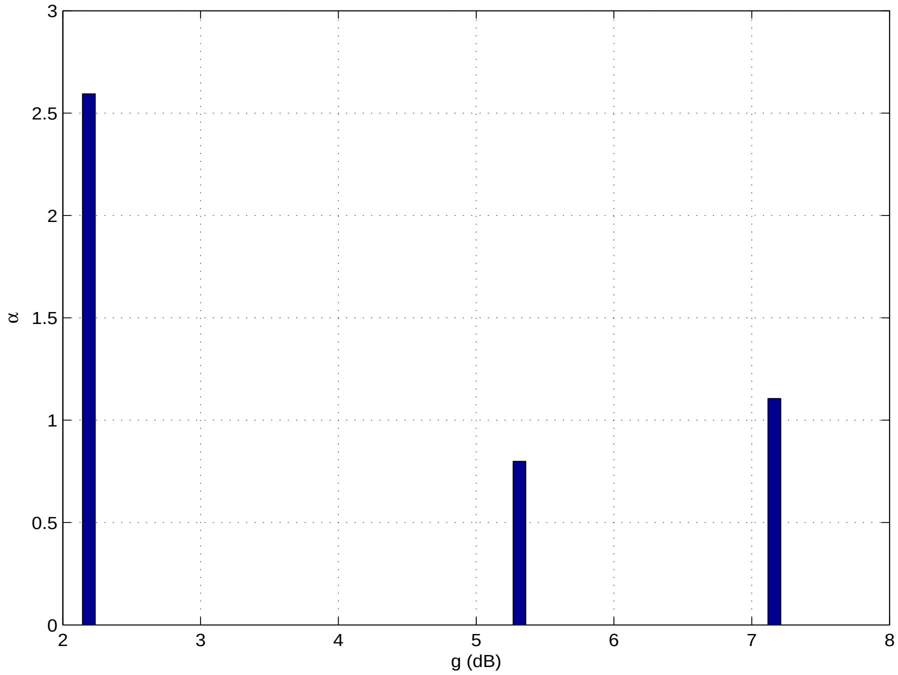

Figure 9: Optimized SNR distribution for the unconditional MMSE-IC receiver with $\alpha = 4.5$, convolutional code with generators $(133, 145, 175)_8$ and $E_b/N_0 = 6.29$ dB.

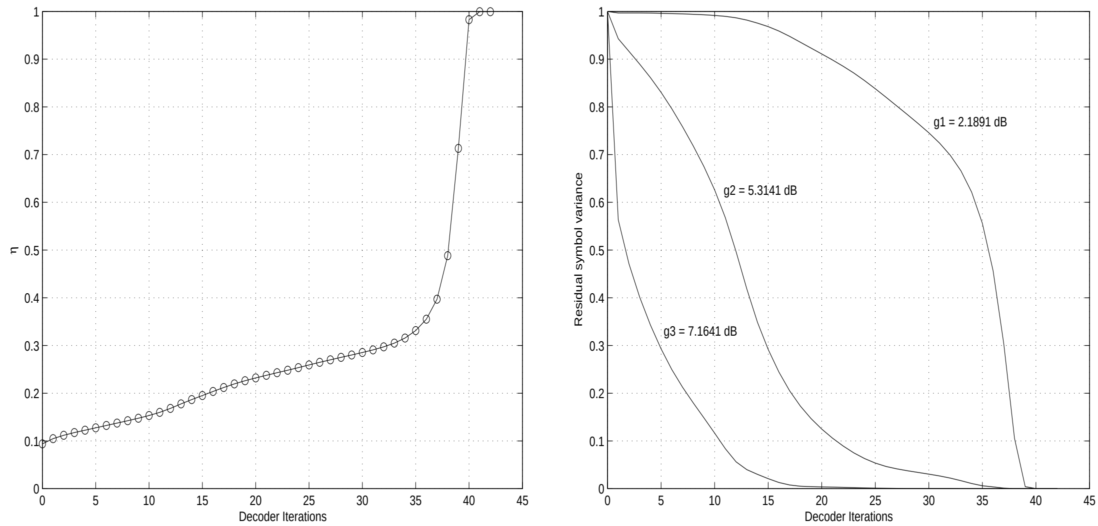

Figure 10: Evolution of the multiuser efficiency (left) and of the user residual symbol variances (right) with the decoder iterations for the system of Fig. 9.

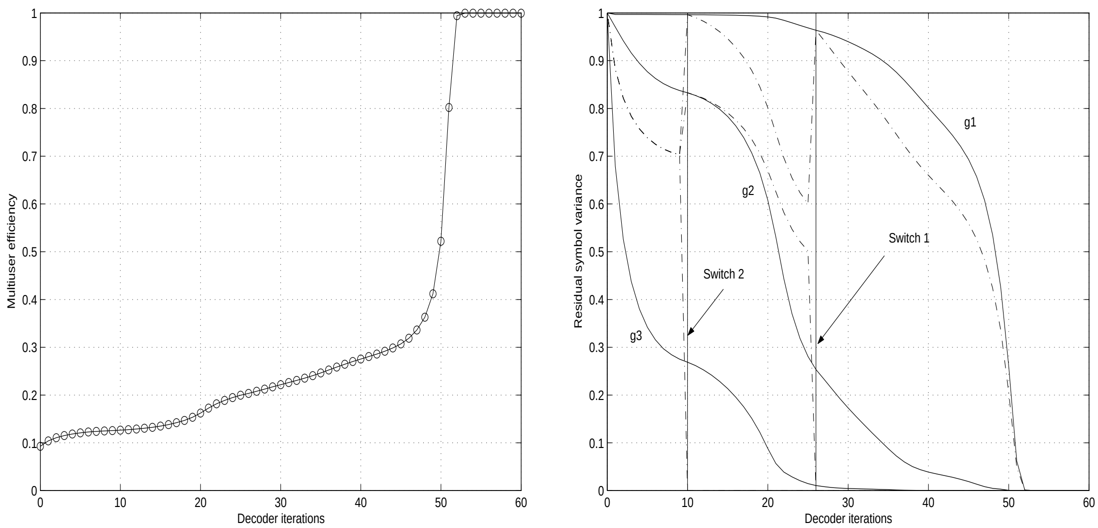

Figure 11: Evolution of the multiuser efficiency (left) and of the user residual symbol variances (right) with the decoder iterations for the system of Fig. 9 with the low-complexity detector algorithm.