# Suboptimal Maximum-Likelihood Multiuser Detection of Synchronous CDMA on Frequency-Selective Multipath Channels

Dan Raphaeli, *Senior Member, IEEE*

***Abstract*—We propose a signal processing technique, based on the estimate-maximize algorithm, in order to perform multiuser code-division multiple-access (CDMA) detection. This algorithm iteratively seeks for the maximum-likelihood solution. The resulting structure is a successive interference cancellation scheme which can be applied to both synchronous and asynchronous CDMA. Higher performance than similar methods is obtained from using deterministic annealing and multiple stages. A soft output is defined, and the signal-to-noise ratio in the soft output of the detector is measured for predicting performance with an outer code with soft input decoder. The new receiver is applied to the problem whereby in a synchronous CDMA system the orthogonality of the codes is destroyed by a frequency-selective channel, caused by multipath fading. This nonlinear technique is shown to perform much better than the minimum mean-square-error linear solution and several other algorithms. The algorithm lends itself to an efficient DSP or VLSI implementation. We evaluate the performance by simulations with coherent quadrature phase-shift keying modulation, known channel and long random Rayleigh multipath. In most cases, we set the number of users equal to the processing gain for maximal throughput. The results are also presented in the form of outage probabilities for random Rayleigh multipath against required fading margin.**

***Index Terms*—Code-division multiple-access, multiuser detection.**

## I. INTRODUCTION

TODAY, code-division multiple-access (CDMA) systems are becoming more and more popular and are moving toward higher rates, especially in the wireless LAN, wireless local loop, and next-generation mobile applications. A typical CDMA system consists of an uplink direction (mobile to base-station) and downlink direction (base-station to mobile). In the uplink each user passes through a different channel, and the users are in most cases asynchronous and thus the codes are nonorthogonal. In a typical downlink, the users are synchronous, the user codes are orthogonal, all users suffer the same channel distortion, and they are coherent with respect to each other. When operating in a high bandwidth, the frequency selectivity of the channel destroys the orthogonality of the downlink as will be explained. Such channels are the indoor microwave and the indoor optical (for $BW \gg 10$ MHz) and the urban (for $BW \gg 1$ MHz). Several authors have considered the multiuser CDMA detection problem in the uplink [1]–[11]. A few have considered it on a multipath channel [2], [6], [11]. Despite the importance of this problem, only few authors investigated the effect of multipath on the transmission in the downlink [12].

In this paper, we address the problem of the downlink with a new powerful nonlinear multiuser detector, and also add results for the linear receiver case. The new method can equivalently be applied to the uplink multiuser problem. Although the detector in this paper was developed for synchronous CDMA, it is straight forward to modify the architecture proposed in Section III-B to asynchronous CDMA.

Such CDMA systems, even with little or no bandwidth expansion, become so powerful in combating the frequency distortion, that they can also be candidates for other applications suffering from multipath. They can nicely compete with multicarrier methods for broadcast applications (or other situations where the transmitter does not know the channel). Like the latter, it requires linear amplifiers at the transmitter and relatively complex circuits for obtaining the high performance. Since a multiuser detector simultaneously detects all users data, such a system becomes very versatile because it allows "bandwidth on demand," dynamic resource allocation with variable data rate, by joining several "users" together. While keeping in mind that the term "user" might be irrelevant to such applications, for the sake of convenience we will use it during this paper.

The nonlinear iterative detector is derived using the estimate-maximize (EM) concept [18]. Multiple-user CDMA detector derived from EM was also independently developed in [19] and in [15, SAGE algorithm]. Note that the resulting detector can also be seen as a nonlinear version of the well known Gauss–Seidel method of linear equation solving. The resulting structure is serial in the sense that it operates on the users one at the time. For each user in its turn, the interference caused by all other users is cancelled using their most updated estimated value. Thus, as each user's estimate is improved, the subsequent user decisions immediately gain from it. This is in contrast to the parallel techniques [2]–[5]. The derivation from the EM algorithms guarantees convergence to a maximum (at least local) of the likelihood metric. In the parallel techniques, convergence is not guaranteed. We tried to implement such a scheme for our problem and observed high probability of divergence when the channel distortion is large. A different

Paper approved by B. Aazhang, the Editor for Spread Spectrum Networks of the IEEE Communications Society. Manuscript received March 21, 1996; revised August 1, 1999. This paper was presented in part at ICUPC'96, Cambridge, MA, September 29–October 2, 1996, and in part at GLOBECOM'96, London, United Kingdom, November 18–22, 1996.

The author is with the Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: danr@eng.tau.ac.il).

Publisher Item Identifier S 0090-6778(00)04013-7.

mathematical representation reveals that the new technique is similar to successive interference cancellation methods like [7] and [8]. The main difference is that in the past only one round of cancellation was used. Here we repeat the operation in multiple rounds (iterations), where in each iteration we successively refine the decision on each of the $K$ users. We note that doing successive interference cancellation in multiple stages (for asynchronous CDMA in the uplink) was also independently suggested by [6], with only *ad hoc* derivation. Another large improvement is obtained here due to the inclusion of deterministic annealing technique [24], [25] to try to avoid local maxima. This essential modification has not been incorporated in [15]. In addition, we use all the known rays of the multipath in order to get the maximal diversity gain. Our formulation of the problem is very simple and can lead to an efficient DSP implementation.

In a typical CDMA system, the base-station to mobile link is synchronous because all the users are transmitted together. Each user out of $K$ users is assigned a code of $N$ chips out of an orthogonal set, such that the cross correlation between users is zero. Then, at the receiver, on a flat (frequency-nonselective) channel, a conventional correlation receiver matched to a particular user will not see any other user interference. However, when the channel is frequency selective, the orthogonality is no longer preserved. The reason is that a codeword $A$ which is orthogonal to a codeword $B$ may not be orthogonal to a phase shift of $B$, which occurs due to multipath. Note that it is not possible to find a set of $K$ codewords which are orthogonal in all shifts, since if it was true then a space of $N$ dimensions would have been spanned by a base of $KN$ vectors (the $K$ codewords each in $N$ possible shifts).

In most of this paper, we consider a fully loaded system so there is no bandwidth expansion. This means that there are $K = N$ co-located users transmitting with the same power, each assigned a code sequence of $N$ chips. It is noted that in a typical application most of the time there will be $K < N$ active users, therefore the result in this paper can be considered as more pessimistic than typical operating conditions.

We assume that the channel impulse response and the carrier phase are known, and in a practical system they can be estimated from a pilot sequence or some other channel estimator [13]. Our tests and the case of [2] indicate that a multiuser detector is not very sensitive to parameters variation.

Mathematically, the detection problem has similarity to the problem of multiuser detection of CDMA with nonorthogonal codes, for example the case of random and independent signatures for each user. This is observed by treating the code+multipath as a new code, and then we can look on the problem as if we have a set of nonorthogonal codes used on a flat channel. As a result, most detectors proposed for the synchronous uplink can be used, and the detector proposed here can also be used for the uplink applications. However, the cross and autocorrelations here behave much differently than in the case of CDMA with random signatures. In particular, the cross correlation becomes much higher on the average, relative to the random case when the channel distortion is high. In addition, here all the users are transmitted together in one signal with equal power and in the same carrier phase. All suffer the same single channel distortion.

The optimal solution to this detection problem, the maximum-likelihood (ML) approach, operates by minimizing the Euclidean distance between the received signal and all $4^K$ possible transmitted signals [for quadrature phase-shift keying (QPSK)]. For a large $K$, the ML approach is not practical, and some suboptimal solution must be employed. We mainly compare our approach with the optimal linear one, the minimum mean square error (MMSE).[1] If the spreading codes are longer than one bit period (as in IS-95 standard for example), it requires a large $\{K \times K\}$ matrix inversion per bit (for cases other than $K = 1$ and $K = N$, see Appendix A) and knowledge of the noise variance. Otherwise the linear approach is computationally much simpler (only $K$ multiplications per user symbol are needed) and can also easily be made adaptive [9]. Furthermore, in the linear case it is not required to detect all the users, but only those required. Once the linear coefficients have been determined, the different users detection can be decoupled, essentially making the linear multiuser detector a bank of single user detectors.

The decorrelating detector is an inferior linear detector (for the completely known parameters case). It is equivalent, in our case, under the model assumptions we use, and for $K = N$, to a zero-forcing equalizer operating at the chip level before a conventional receiver. Indeed, it shows poor performance (5–7 dB worse than MMSE for random channels) due to large noise enhancement effects. We provide a comparison to other suitable algorithms: the recent parallel interference cancellation presented in [5], MC-CDMA [22], the decision feedback [23], and SAGE [15]. Finally, we provide the result of coded orthogonal frequency-division multiplexing (OFDM) simulated on the channels used as examples.

The paper is organized as follows. First, we present the system model. Next, we derive the iterative detector and show some equivalent structures. Generation of soft output for coded systems is suggested in Section IV together with an effective SNR model. Some simulation results are given in Section V. We summarize with conclusions.

## II. System Model

The mathematical formulation of the system is as follows (we are considering the CDMA downlink). At each symbol period $T, 2K$ bits of data (either uncoded or coded bits) are transmitted. The data is used to modulate $K$ QPSK symbols $u_k = \pm 1 + \pm j$. Each $u_k$ belongs to one "user." As mentioned, it may not represent a true user data. Each user $k$ is assigned a unique code $\mathbf{c}_k$ of length $N$, normalized to unit energy, which is multiplied by the user data $u_k$. Unless stated otherwise, we set $K = N$ for full bandwidth use. The codes should, at least approximately, be an orthonormal signal set. A good choice for a set of codewords which we are using, is a set of Gold codes, chosen such that the normalized cross correlation without shifts is $1/N$ (close to orthogonal), and the shifted cross correlations are bounded and behave pseudorandomly. No significant performance difference was experienced when another randomly selected orthonormal code was used.

[1] The MMSE is optimal in the sense that the average signal-to-noise ratio (SNR) is maximized. The BER is only approximately minimized.

Let $\mathbf{c}_k$ be the $k$th column of the $N \times K$ code matrix $C$. Each user symbol is taken from QPSK constellation $u_k = \pm 1 + \pm j$. The users are summed together to form the output signal

$$x_i = \sum_{k=0}^{K-1} u_k c_{i,k}, \qquad i = 0 \cdots N-1. \tag{1}$$

The signal $x_i$ is sent over an equivalent discrete channel (which include the effect of pulse shaping and analog filtering) [16, p. 588][2] where its output $r_i$ can be represented as

$$r_i = y_i + n_i = \sum_{j=0}^{L} h_j x_{i-j} + n_i \tag{2}$$

where $h_j$ are the channel impulse response coefficients and $n_i$ is a white Gaussian noise. We assume that $N > L$. Apart from the cochannel interference, we also have ISI between successive data symbols going through the channel. This ISI can be largely cancelled by using decision feedback, i.e., taking the previously detected user symbols, computing their contribution to the current symbol waveform, and subtracting the same. Another option is to include a guard band, or just to ignore the interference if $N \gg L$.

If decision feedback is used, the ratio between the energy of one bit of one user to the average energy of a residual error due to $b$ detection errors in previous symbols is

$$\frac{E_b}{I} = \frac{N^2}{4bL_e} \tag{3}$$

where $L_e$ is related to the delay spread of the channel, see Appendix B. This residual error can be neglected for common values of $N$ which are in the order of 100. Therefore, we will neglect it in the remainder of this paper.

The channel can be expressed in a matrix form by the $N \times N$ channel matrix

$$H_{j,k} = \begin{Bmatrix} h_{j-k}, & \text{if } 0 \le j-k \le L \\ 0, & \text{otherwise} \end{Bmatrix}. \tag{4}$$

Define the vectors $\mathbf{r} = \{r_0, \cdots, r_{N-1}\}^T$ and $\mathbf{n} = \{n_0, \cdots, n_{N-1}\}^T$, then

$$\mathbf{r} = S\mathbf{u} + \mathbf{n} \tag{5}$$

where $S = HC$. The $k$th column of $S$, $\mathbf{s}_k$, is the distorted code of user $k$ at the channel output, truncated to $i = 0 \cdots N-1$. Equation (5) can be also expressed in the form

$$\mathbf{r} = \sum_{k=0}^{K-1} u_k \mathbf{s}_k + \mathbf{n}. \tag{6}$$

Note that the (5) or (6) are also appropriate for describing the synchronous CDMA *uplink*.

## III. Iterative Detector Development

In this section, we develop a type of multiuser detector for CDMA, here applied for the downlink channel, but is equally applicable to the uplink channel (both synchronous and asynchronous cases) with few modifications. For the asynchronous case one should use the implementation described in Section III-B with some simple modifications. Using (6) the problem in hand suits the model of superimposed signal in white noise treated in [18]. Following this reference, the EM algorithm is reduced to the following iterative process:

$$u_k^{(n+1)} = \text{csign}\left[\mathbf{s}_k^\dagger \mathbf{s}_k u_k^{(n)} + \beta_k^{(n)} \left(\mathbf{s}_k^\dagger \mathbf{r} - \sum_{l=0}^{K-1} \mathbf{s}_k^\dagger \mathbf{s}_l u_l^{(n)}\right)\right] \tag{7}$$

where $\text{csign}(a+jb) = \text{sign}(a) + j\text{sign}(b)$, $\beta_k^{(n)}$ are arbitrary chosen coefficients satisfying $\sum_{k=0}^{K-1} \beta_k^{(n)} = 1$, and $u_k^{(n)}$ is the $n$th iterative estimate of $u_k$. A particular choice of $\beta_k^{(n)}$ is more easy to implement and fast converging, and is $\beta_k^{(n)} = \delta(n - k \bmod K)$, i.e., a unit in the $n$th location and zero in the rest of the locations. With this choice, exactly one user is updated in each iteration. It thus takes $K$ iterations to update each user once.

Since the EM guarantees that the likelihood will be increased in each step, convergence to a maximum will be obtained. However, local maxima exists and should be avoided as much as possible. Thus, we added the deterministic annealing technique to the above basic approach. The main method in deterministic annealing is replacing the decision function (sign) with a soft decision function, with the amount of "softness" controlled by the a parameter called "temperature" in analogous to crystallization process. As the iterations proceed, the temperature is decreased, which ultimately forces hard decisions. We have chosen the hyperbolic tangent function for implementing the soft decision. If the modulation is nonbinary, e.g., QAM, the soft decision function is obtained by using a smoothed staircase function for the decision function.

The algorithm can be also expressed in a form more suitable for implementation. Define a $K$ elements array $\{\hat{u}_k\}$ which through the algorithm will always contain the best estimation of $\mathbf{u}$. For the mathematical notation, we will add the index $i$ to denote the iteration number. Each iteration is equivalent to $K$ iterations of (7). The algorithm is as follows. At the initialization step, compute the outputs of bank of filters matched to $\mathbf{s}_k$

$$Y_k = \mathbf{s}_k^\dagger \mathbf{r}, \qquad k = 0 \cdots K-1 \tag{8}$$

and the cross-correlation matrix $G$ which its elements are

$$g_{k,l} = \mathbf{s}_k^\dagger \mathbf{s}_l, \qquad l = 0 \cdots K-1;\ k = 0 \cdots K-1. \tag{9}$$

The vector $Y$ can be the output of $K$ RAKE receivers, but it is more efficient to have a single filter matched to the channel and the chip pulse shaping, followed by a bank of correlators which correlate with $\mathbf{c}_k$. Initialize the estimated array as $\hat{u}_{k,0} = Y_k \quad k = 0 \cdots K-1$ (this is the best estimate available at this point in the algorithm). With bad channels, $Y_k$ is so corrupted that initialization with zeros or by random values brings about the same performance after a few iterations. If the complexity of a MMSE receiver can be paid, initialization by the MMSE receiver output speeds up convergence considerably.

At each iteration $i$ of the algorithm, the users are processed one by one by letting $k$ go from 0 to $K-1$. This order is

[2]The whitened matched filter model is used only for the easy mathematical treatment and is not necessary for actual receiver implementation. Furthermore, it is not relevant if the channel is minimum phase or not.

arbitrary. In situations of unequal power users, e.g., in the uplink, a preferred order to speed up the convergence is the order of decreasing SNR at the matched filters output. For each user $k$, compute

$$\hat{u}_{k,i} = f(Y_k - \hat{I}_{k,i}) \tag{10}$$

where

$$\hat{I}_{k,i} = \sum_{l=0}^{k-1} g_{k,l}\hat{u}_{l,i} + \sum_{l=k+1}^{K-1} g_{k,l}\hat{u}_{l,i-1} \tag{11}$$

and

$$f(x) = \tanh\left(\frac{\mathrm{Re}(x)}{2T_i}\right) + j\,\tanh\left(\frac{\mathrm{Im}(x)}{2T_i}\right) \tag{12}$$

is a soft decision function (separately on the real and the imaginary parts), performing the deterministic annealing, where $T_i$ is the temperature, i.e., a parameter controlling the softness. The value $\hat{I}_{k,i}$ is essentially the total interference from all the other users on user $k$. For most channels checked, 6 iterations were sufficient (only two to three iterations are needed if initialization by MMSE is made). Only a few very bad channels could benefit slower annealing and more iterations. Using this choice, the number of multiplications per user symbol becomes $6K$ (for fix channels, without the computation of the cross correlations). The above algorithm is applicable equally well for the synchronous uplink or the downlink, but from this point in the paper the derivations and results are applicable to the downlink only.

### *A. Reduced Complexity Direct Implementation of the Algorithm*

A fast computation of $G$ can be obtained as follows. Let $\gamma_{p,l,m}$ be the $(l, m)$th element of the $K \times K$ matrix $\Gamma_p$ and is given by

$$\gamma_{p,l,m} = \sum_{i=0}^{N-1} c_{i,k}^* c_{i+p,k}. \tag{13}$$

If the codes are fixed, then $\Gamma_p$ are precomputed and stored in memory. Then, after some algebra

$$G = \sum_{p=-L}^{L} \left( \sum_{q=\max(0,-p)}^{\min(L, L-p)} h_q^* h_{q+p} \right) \Gamma_p. \tag{14}$$

Note that computations of $G$ are also needed for the MMSE linear detector, if used (for $K < N$). Thus, additional $(2L+1)K^2$ multiplications are needed for the computation of $G$ for a fixed code each time the channel is changed.

A specific code structure allows both fast computation of $G$ and $Y$. Assume $K = N$. Let $W$ be a Hadamard matrix (or other orthonormal matrix having fast multiplication algorithm) and $v_i$ a random binary code of length $N$ (or more generally complex numbers of random phase and unit magnitude). Then, the codewords $\mathbf{c}_k$ are obtained by multiplying the columns of $W$ with $v_i$ component-wise, and are clearly orthonormal. The randomness of the codewords which are achieved through the multiplication by $v_i$ is useful to avoid sustained worst case conditions and for separating adjacent cells. Define $V$ as the diagonal matrix having $v_i$ on its diagonal, then $C = VW$. Now we obtain

$$G = W^\dagger V^\dagger H^\dagger H V W, \quad Y = S^\dagger \mathbf{r} = W^\dagger V^\dagger H^\dagger \mathbf{r}. \tag{15}$$

It can be shown that both $G$ and $Y$ can be efficiently computed by the fast Hadamard transform and few additional operations.

### *B. Indirect Implementation Using Multiple-Stage Successive Interference Cancellation*

An alternative implementation to the iterative algorithm results in an architecture of a multiple-stage successive interference cancellation (MSSIC). Here, there is no need to explicitly compute neither (8) nor (9). The initialization of $\mathbf{u}_{k,0}$ will be with zeros. We start from (10).

$$\begin{aligned}\hat{u}_{k,i} &= f(Y_k - \hat{I}_{k,i}) \\ &= f\left[\mathbf{s}_k^\dagger \left(\mathbf{r} - \sum_{l=0}^{k-1} \mathbf{s}_l \hat{u}_{l,i} + \sum_{l=k+1}^{K-1} \mathbf{s}_l \hat{u}_{l,i-1}\right)\right] \\ &= f(\mathbf{s}_k^\dagger \mathbf{z}_{k,i})\end{aligned} \tag{16}$$

where

$$\mathbf{z}_{k,i} = \mathbf{r} - \sum_{l=0}^{k-1} \mathbf{s}_l \hat{u}_{l,i} + \sum_{l=k+1}^{K-1} \mathbf{s}_l \hat{u}_{l,i-1}. \tag{17}$$

The meaning of $\mathbf{z}_{k,i}$ is the user $k$ with all known interference removed. A recursion equation for $\mathbf{z}_{k,i}$ can be obtained

$$\mathbf{z}_{0,0} = \mathbf{r} \tag{18}$$

$$\mathbf{z}_{k,i} = \begin{cases} \mathbf{z}_{k-1,i} - \mathbf{s}_{k-1}\hat{u}_{k-1,i}, \\ \quad \text{if } i = 0 \text{ and } 0 < k < K \\ \mathbf{z}_{k-1,i} - \mathbf{s}_{k-1}\hat{u}_{k-1,i} + \mathbf{s}_k \hat{u}_{k,i-1}, \\ \quad \text{if } i > 0 \text{ and } 0 < k < K \\ \mathbf{z}_{K-1,i-1} - \mathbf{s}_{K-1}\hat{u}_{K-1,i-1} + \mathbf{s}_0 \hat{u}_{0,i-1}, \\ \quad \text{if } i > 0 \text{ and } k = 0 \end{cases}. \tag{19}$$

The operation of the MSSIC is very intuitive. There is a work signal $z$ which is initiated by the received signal. Suppose we are at the point at iteration $i$ just after a user $k-1$ was estimated. First, it is remodulated and removed from the work signal for reducing interference for the next users. Then, we move to the next user estimation, user $k$ (user 0 if the previous was $N-1$). This user was removed last time it was estimated (in iteration $i-1$), so it is remodulated and added back to the work signal since its energy is needed for its own decision. After the decision, it will be removed again for not interfering with the other users.

We like to further simplify the algorithm toward actual implementation. Let $\mathbf{z}'_{k,i} = H^\dagger \mathbf{z}_{k,i}$, and $\mathbf{s}'_k = H^\dagger \mathbf{s}_k = H^\dagger H \mathbf{c}_k$. Then, the algorithm becomes

$$\hat{u}_{k,i} = f(\mathbf{c}_k^\dagger \mathbf{z}'_{k,i}) \tag{20}$$

$$\mathbf{z}_{0,0} = H^\dagger \mathbf{r} \tag{21}$$

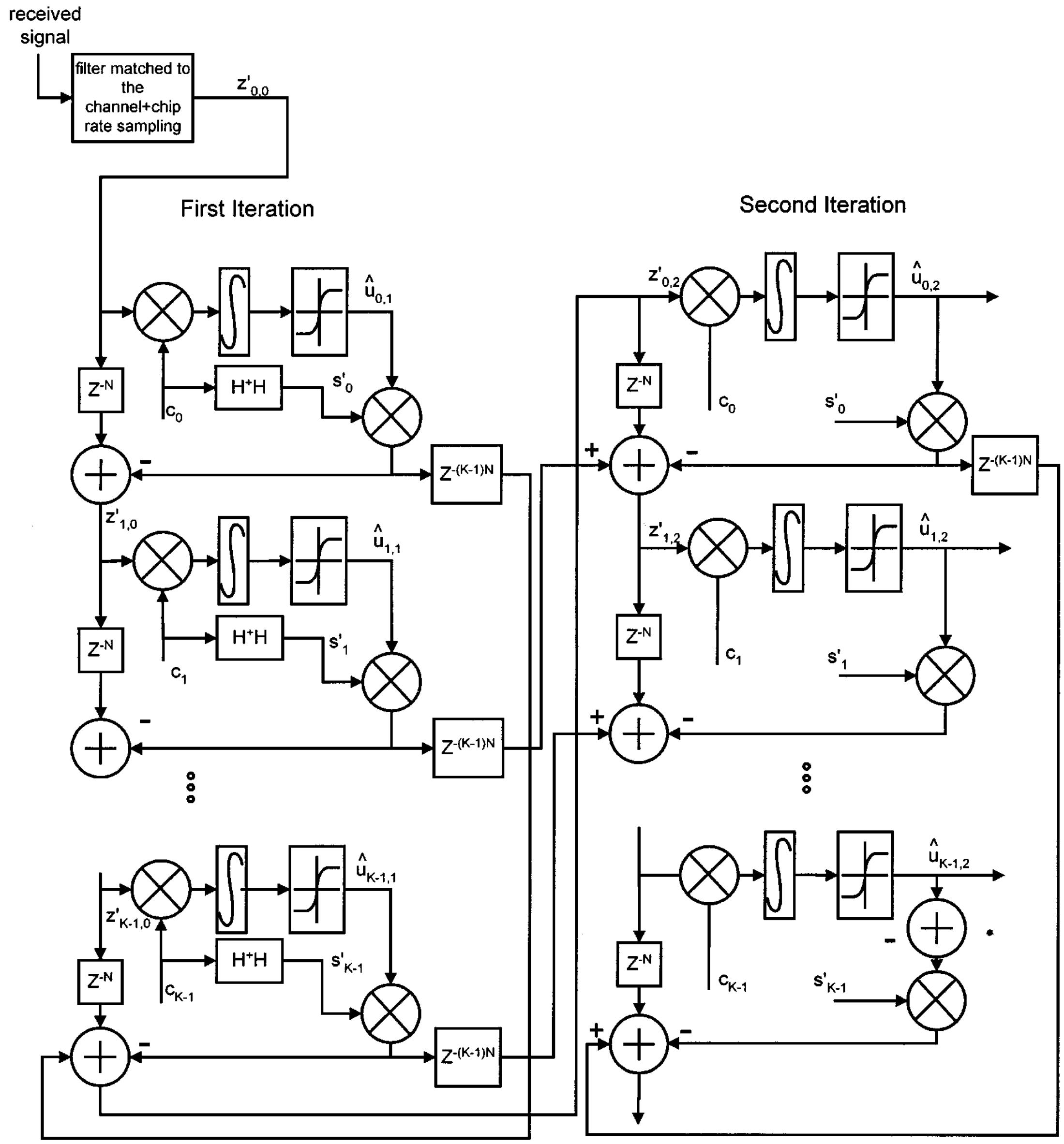



Fig. 1. MMSIC.

$$\mathbf{z}'_{k,\,i} = \left\{ \begin{array}{l} \mathbf{z}'_{k-1,\,i} - \mathbf{s}'_{k-1}\hat{u}_{k-1,\,i}, \\ \quad \text{if } i = 0 \text{ and } 0 < k < K \\ \mathbf{z}'_{k-1,\,i} - \mathbf{s}'_{k-1}\hat{u}_{k-1,\,i} + \mathbf{s}'_{k}\hat{u}_{k,\,i-1}, \\ \quad \text{if } i > 0 \text{ and } 0 < k < K \\ \mathbf{z}'_{K-1,\,i-1} - \mathbf{s}'_{K-1}\hat{u}_{K-1,\,i-1} + \mathbf{s}'_{0}\hat{u}_{0,\,i-1}, \\ \quad \text{if } i > 0 \text{ and } k = 0 \end{array} \right\}. \quad (22)$$

The implementation block diagram of the MSSIC is shown in Fig. 1 (we ignore for simplicity the decision feedback described in Section II which may also be required). The operation (21) translates in the implementation to a matched filter (usually implemented by oversampled finite-impulse response) followed by a chip rate sampling. The block $H^{\dagger}H$ in the figure which implements $H^{\dagger}H\mathbf{c}_k$ is a finite-impulse response filter whose impulse response is the convolution of $h_i$ with $h^{*}_{-i}$. The computation effort for each stage needed here is larger than the previous approach. This balance may reverse if we take into account the computation of $G$ if the code changes for each symbol transmitted. It is to be noted that the MSSIC approach seems more appealing for VLSI implementation.

## IV. Coded Systems

In a conventional CDMA system, each user needs to have a separate error correcting encoder and decoder. Here, it is possible to employ only a single common encoder and decoder for the whole system, since both the transmitter and the receiver have all the symbols of all the users available. In this case, there is only one encoder at the base station instead of $K$ encoders, but on the other hand, the decoder at the mobile is working $K$ times faster, if only one user data is used at that terminal. An additional advantage to the system is that the decoding delay is reduced by $K$. It is desired also to use an interleaver in order to make the users decisions independent for optimizing the performance with an error correcting code which is designed for a memoryless channel. We recommend a short bit-level block

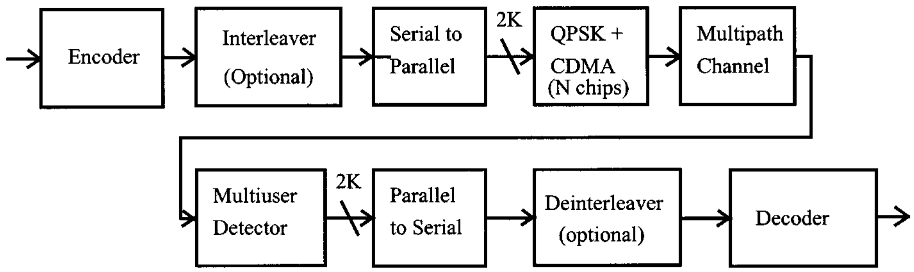



Fig. 2. A coded point to multipoint CDMA data transmission system.

interleaving ($\mathrm{depth} \times N$). Omitting the interleaver cause some additional 1–2-dB loss (experimentally). Note, that in the case there is a separate decoder for each user, the results of this section apply if there is a random permutation of users codes, or approximately apply otherwise. The interleaver in this case is obviously unnecessary.

The best performance could be achieved by including the decoder in the iterations. This means an algorithm based on turbo-decoder principle that alternately operate the interference cancellation and the decoder. For saving complexity, memory requirements and delay or fitting into a system where the decoder already exists, we choose not to use this approach. We assume that the detector works on the uncoded data, and after the last iteration it provides soft output to the decoder. The complete system is shown in Fig. 2.

We use a simple technique to produce soft output for the decoder. We simply output the term $\tilde{u}_k = Y_k - \hat{I}_k$ just before the decision device at the last stage. Ideally, with perfect cancellation, this term contains only the symbol of user $k$ with additive thermal noise, the same as the symbol transmitted over AWGN. In this case, it is a perfect soft output.

The outer code sees an equivalent channel its input being the interleaver input and its output being the deinterleaved soft output. If we assume that this channel is additive Gaussian, we can specify the effective SNR for this channel and then one can evaluate the performance of the system with any outer code. Practically, the residual interference makes the soft output less conditionally Gaussian. This causes an additional 0.5–1-dB degradation above the model, as indicated in the results. The linear MMSE detector is also modeled in the same way. In this case the Gaussian model is much more accurate.

Under the Gaussian assumption, the soft output is modeled as

$$\tilde{u}_i = \alpha_{k(i)} u_i + w_i \tag{23}$$

where $\alpha_{k(i)}$ are deterministic values, and $w_i$ are independent complex Gaussian r.v. with variance $\sigma^2_{w,\,k(i)}$. Both $\alpha_k$ and $\sigma^2_{w,\,k}$ are functions of the user $k$, which can be determined from $i$ using the deinterleaver input to output relation $k(i)$.

The parameters $\alpha_k$ can be found by correlating $\tilde{u}_i$ and $u_i$ and is unity for a flat channel and ideal cancellation. Using this model the effective SNR (averaged over the different users) becomes $\overline{\alpha_k^2/\sigma^2_{w,\,k}}$. Both parameters have been statistically measured by simulation of the iterative algorithm and the MMSE algorithm and serve for performance measure. Consequently, our results can be useful to evaluate performance with any outer code one chooses to use. The effective $E_s/N_0$[3] also matches the uncoded BER at low and moderate $E_s/N_0$. When the uncoded BER becomes less than $10^{-4}$, local maxima trapping start to dominate performance, and error floor is formed. This is not a problem when coding is applied.

Assuming a pseudorandom interleaver, the distribution of $\alpha_k$, results in a situation very similar to a fading channel, where $\alpha_{k(i)}$ is approximately a stationary random variable. The fading is accounted for by computing the Chernoff factor [20, Appendix A]

$$\begin{aligned} C(s,\,\hat{s}) &= E\left[\exp\left(-\frac{\alpha^2_{k(i)}}{4\sigma_{w,\,k(i)}}|s-\hat{s}|^2\right)\right] \\ &= \frac{1}{K}\sum_{k=0}^{K-1} \exp\left(-\frac{\alpha_k^2}{4\sigma_{w,\,k}}|s-\hat{s}|^2\right). \end{aligned} \tag{24}$$

In our case, the expectation will be carried out first over the random interleaver and then over the $K$ users. The symbols at the channel input are $\pm 1$ (assuming that the interleaving is performed in the bit level rather than in the QPSK symbol level), hence $|s-\hat{s}|^2 = 4$. A degradation factor relative to constant effective SNR is then computed by

$$\Lambda = -\ln\left(\frac{\frac{1}{K}\sum_{k=0}^{K-1}\exp\left(-\frac{\alpha_k^2}{\sigma_{w,\,k}}\right)}{\overline{\alpha_k^2/\sigma^2_{w,\,k}}}\right). \tag{25}$$

An approximation to $\alpha_k$ is obtained as follows:

$$\begin{aligned} \alpha_k &= \frac{1}{2}E[(Y_k - \hat{I}_k)u_k^*] \\ &= \frac{1}{2}E\left[\left(u_k g_{kk} + \sum_{\substack{l=0\\ l\neq k}}^{K-1} g_{k,\,l}u_l - \sum_{\substack{l=0\\ l\neq k}}^{K-1} g_{k,\,l}\hat{u}_l + \mathbf{s}_k^\dagger \mathbf{n}\right)u_k^*\right] \\ &= g_{kk} - \frac{1}{2}\sum_{\substack{l=0\\ l\neq k}}^{K-1} g_{k,\,l}E[\hat{u}_l u_k^*]. \end{aligned} \tag{26}$$

In the last term, the correlation between $\hat{u}_l$ and $u_k$ is averaged on all users, and the result can be neglected, leaving $\alpha_k = g_{kk}$.

Following (26), $w_k$ is the sum of two independent terms

$$w_k = \sum_{\substack{l=0\\ l\neq k}}^{K-1} g_{k,\,l}(u_l - \hat{u}_l) + \mathbf{s}_k^\dagger \mathbf{n}. \tag{27}$$

[3] $E_s$ denotes the energy of a coded bit, i.e., one coordinate of $u_i$.

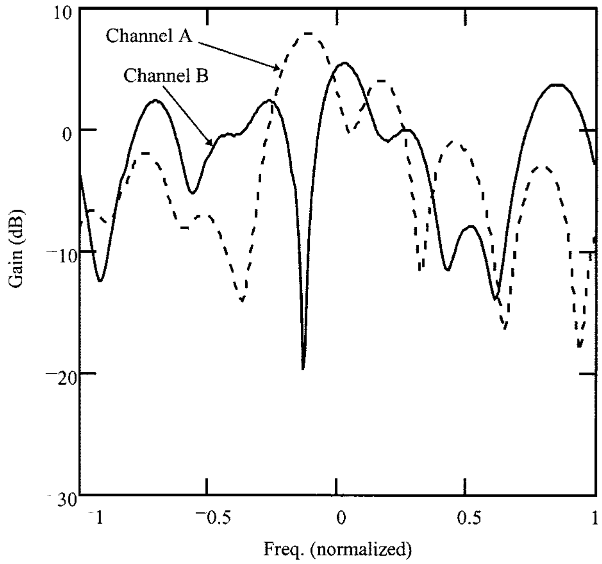



Fig. 3. The spectrum of two example channels.

The last term, the thermal noise contribution, has variance $g_{k,k}\sigma^2$. The overall noise contribution is difficult to analyze since the user error terms $u_l - \hat{u}_l$ are not independent, and moreover, they are dependent on $n_k$. However, in most cases, when the degradation is low, the thermal noise term is significantly larger. In other cases, we have observed that the overall noise is still a strong function of $g_{k,k}$. Ideally, the decoder should be informed with the time-varying noise variance. However, if the noise variance is proportional to $g_{k,k}$ then from (23) we see that the soft value is already in the correct form to be used without any modification in the branch metric calculation of a Viterbi outer decoder.

## V. Simulation Results

### A. Results of the Proposed Algorithm

We have simulated 5000 randomly selected Rayleigh faded channels of ten taps. We have used a set of Gold codes of length $N = 63$. No ISI between symbols was assumed as explained in Section II. The channels were generated by having each tap be an independent complex Gaussian random variable. The deterministic annealing formula was chosen empirically and optimized for six iterations as $T_i = 0.69 \cdot 1.3^{-i}$. Some bad channels could gain from having slower temperature decrease and a larger number of iterations. Unless mentioned, we used $K = N$ to represent the fully loaded condition of the system in term of bandwidth use. Note that in the results the effective SNR are always reduced by the Chernoff factor. The effective SNR is an average over the different users. The SNR as a function of $k$ is random-like and is about 3 dB peak-to-peak for typical channels and 6 dB for bad channels.

The spectrum of two channel examples is shown in Fig. 3. Channel A is one of the worst channels in the channels set. Channel B is just a randomly picked channel (number 100 in the set). The channels were normalized such that there is a unit power gain for input with flat spectrum. The effective SNR after the last iteration of the algorithm and at various stages is shown, and compared to the MMSE effective SNR and to the ideal case

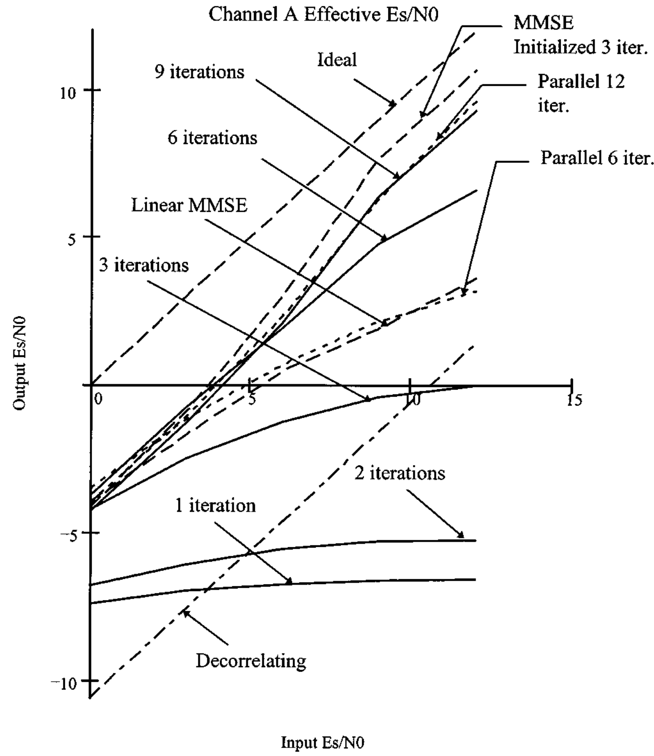



Fig. 4. Effective SNR of channel A ($K = N = 63$).

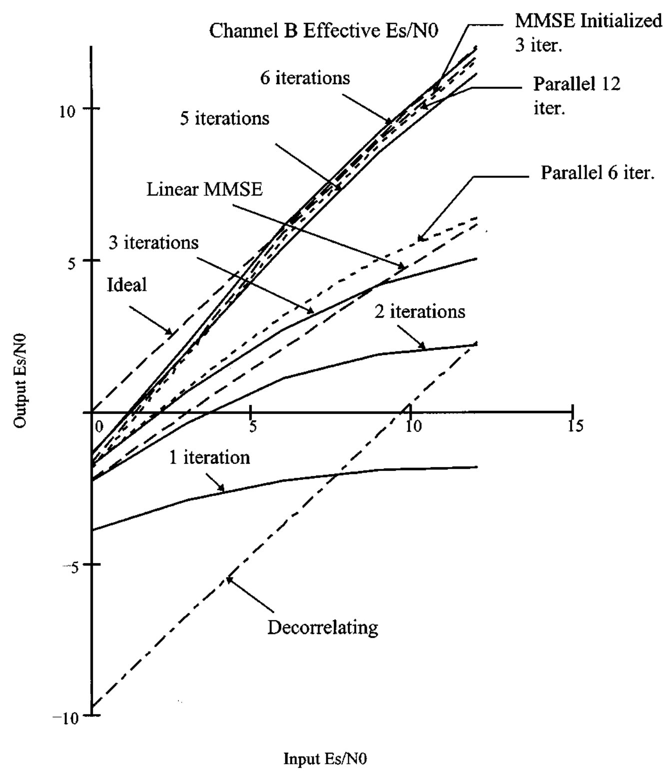



Fig. 5. Effective SNR of channel B ($K = N = 63$).

occurring on a flat channel in which there is no SNR loss in the equivalent channel. The results are shown in Figs. 4 and 5. Both the MMSE and the nonlinear receiver suffer a large degradation with channel A, but there is a noticeable advantage with the nonlinear one. As mentioned above, having slower temperature

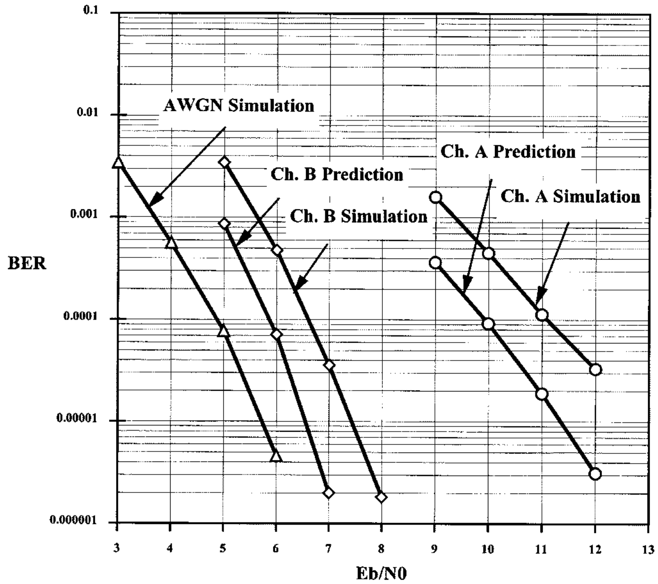



Fig. 6. Simulation with four-state rate 1/2 convolutional code (6 iterations detector, $K = N = 63$).

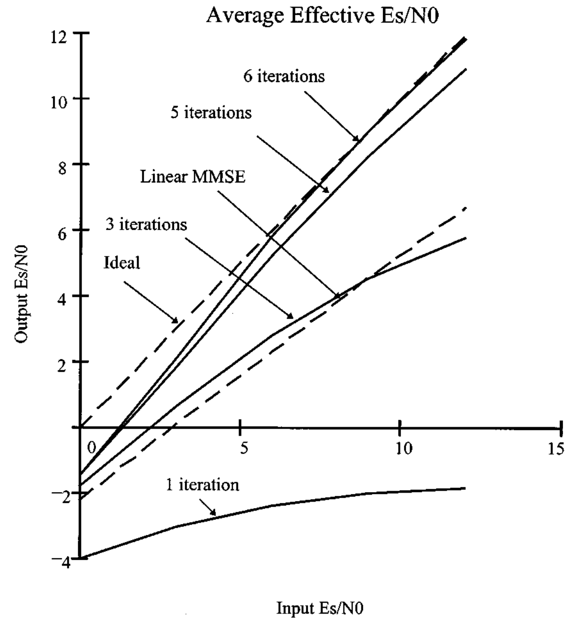



Fig. 7. Average effective SNR for ten constant profile Rayleigh faded taps ($K = N = 63$).

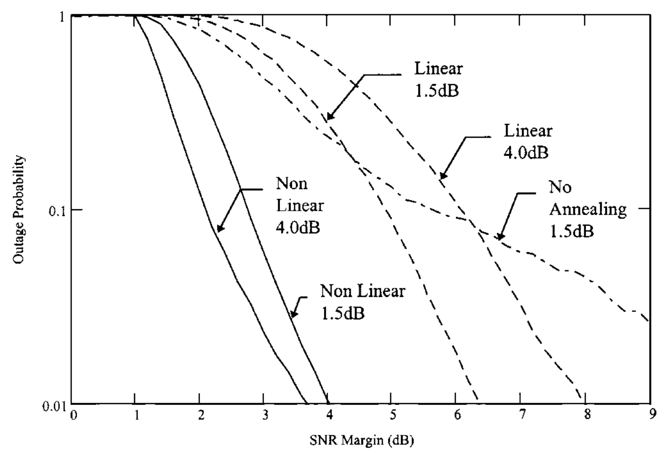



Fig. 8. Outage probability for ten constant profile Rayleigh faded taps ($K = N = 63$). The margin is defined as the amount of increase in transmitted power relative to the nominal one, and outage happens when the output effective SNR is lower than the nominal value.

decrease and a larger number of iterations would improve the performance further in 1–2 dB. In the typical channel, channel B, we observe that the nonlinear algorithm performs close to ideal after six iterations, while the linear detector suffers from more than 4-dB degradation. We also show the performance of the decorrelating detector. Note that as shown in Appendix A, for $K = N$, the decorrelating detector is equivalent to a zero-forcing equalizer in front of a conventional receiver. Please refer to the introduction for a complexity comparison between the linear and nonlinear approaches. A combination of the above two approaches has the best performance. If the complexity of a MMSE receiver can be paid, initialization of $\hat{u}_{k,0}$ by the MMSE receiver output speeds up convergence considerably and result in an improved performance. The performance after three iterations of the nonlinear algorithm after initialization by the MMSE receiver is also shown in Figs. 4 and 5.

Fig. 6 shows the scheme performance with four states rate 1/2 convolutional code as an outer code and the two specific channels A and B. We have used a short bit-level block interleaving ($10 \times 2K$) to cause correlated bits to be separated by at least 9 at the decoder. The prediction curves were produced by having the effective SNR from the uncoded simulation as input to the code performance curve as simulated on a flat AWGN channel. The difference between the predicted and actual performance is due to the non-Gaussian distribution of the soft output, and amounts to about 1 dB.

The average effective SNR results of the channels set is shown in Fig. 7. The average performance is most influenced by the performance on typical channels like channel B. We observe almost no degradation after six iterations of the algorithm, while the MMSE suffers more than 4 dB.

For some applications, the outage probability has much more practical significance than the average performance. The outage probability is the probability that a channel picked at random causes communication failure. In designing a system expected to suffer multipath effects, one increases the power by a fading margin amount such that the outage probability is less than certain limit. There are two factors involved in the outage probability: the fading of the signal energy and the spectral distortion. In the case we simulated, the diversity of the ten taps makes the probability of significant fade negligible, relative to the effects of the spectral distortion.

A loss of 1 dB was added to the nonlinear effective SNR for accounting the soft-output loss. The results are shown in Fig. 8 for nonlinear and linear detectors at two working points. One is $E_s/N_0 = 1.5$ dB which is the threshold for a standard rate 1/2, $K = 7$ convolutional code at $P_b = 10^{-5}$ and the second is $E_s/N_0 = 4.0$ dB which is the threshold for the rate 3/4, $K = 7$ convolutional code. As the SNR required is higher, the linear case suffers more from the noise enhancement effect. At the same time the nonlinear performs even better at high SNR as

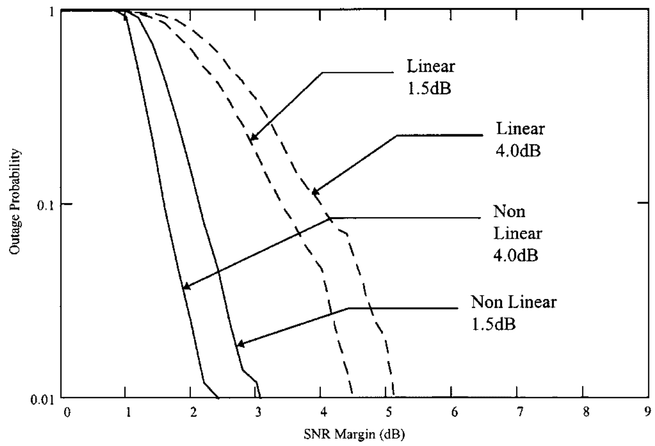



Fig. 9. Outage probability for ten constant profile Rayleigh faded taps ($K = 48, N = 63$).

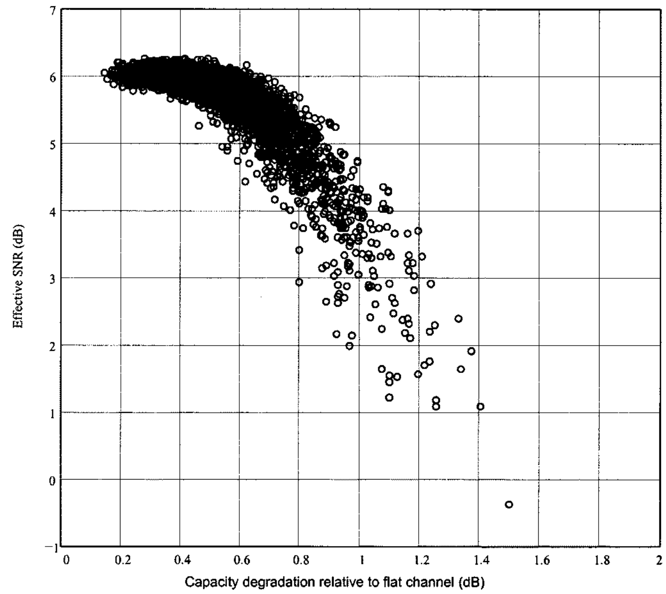



Fig. 10. Effective SNR versus capacity degradation at rate 1/2 and $E_s/N_0 = 6$ dB.

the tentative decisions improve. We can see 2.4- and 4.4-dB differences, respectively, at the two working points between linear and nonlinear detectors for 1% outage. If one uses uncoded information the working point for $P_b = 10^{-5}$ is around 10 dB (not on the graph) and the linear detector will require 19-dB margin!

In some systems, there is a tradeoff between the bandwidth dedicated to the coding redundancy and the bandwidth wasted by a reduction of $K$. Both provides performance gain. For example, we can use a rate 1/2 code and $K = N$ or rate 2/3 code and $K/N = 3/4$. Both alternatives give the same data rate. The reduction of the fading margin indicated in Fig. 9, is to be compared to the coding gain loss moving from rate 1/2 to a rate 2/3 code.

The capacity (for flat transmitted power distribution, the channel is unknown at the transmitter) is decreased as the channel flatness decreases. It is reasonable that our system will behave similarly. As intuitively expected, the degradation of our system is higher than the capacity degradation. The capacity degradation of a channel means the increase in minimum $E_s/N_0$ relative to that needed with flat channel, both with the same given code rate. The minimum $E_s/N_0$ for transmitting at the capacity of a frequency-selective channel with a flat transmitted power is given by solving [16]

$$\int \log_2\left(1 + \frac{E_s}{N_0 W}|H(f)|^2\right) df = r \tag{28}$$

where $r$ is the code rate. The results for $r = 0.5$ are shown in Fig. 10. The results of the effective SNR are given after the sixth iteration at received SNR of $E_s/N_0 = 6$ dB (of coded bits).

### *B. Comparison to other Algorithms*

A comparison with other suitable algorithms were made. The SAGE algorithm [15] is identical to our algorithm, but without the deterministic annealing. The result of the simulation of the algorithm with hard decisions for channels A and B were not added to Figs. 4 and 5 since they are already too crowded. For channel A the effective SNR was below $-1$ dB for all input SNR. The large degradation is due to the poor starting point (matched filter output) which causes the algorithm most frequently to reach a local minimum. The situation with channel B is different. The results are similar to the results obtained with the annealing. In this case the starting point is typically good enough to cause the algorithm to reach global minimum. About 10% of the channels cause poor convergence, as shown in Fig. 8.

One alternative nonlinear detector is the multistage parallel interference cancellation [5]. For a severely distorting channel the algorithm fails to converge ($\mathrm{BER} \cong 0.5$) unless the step size parameter $p_i$ is reduced. However, this reduction causes slow convergence. By empirical tries we found that for channel A $p_i = 0.4$ is maximum. The soft output and hence the effective SNR can be found by the same equations as for the MSSIC. The results are shown in Figs. 4 and 5. We can conclude that the parallel approach requires about twice the number of iterations and has some probability to diverge.

MC-CDMA has been considered a candidate to combat multipath [22]. Consider the matrix formed by $C' = C\Omega$, where $\Omega$ is the $N \times N$ DFT matrix. $C'$ will have the same statistical properties of a random orthogonal matrix. Thus, MC-CDMA is identical to the CDMA considered here. The controlled equalization considered in [22] is identical to a decorrelating receiver if the threshold is 0, and approximate the MMSE solution for optimized threshold. The MMSE results shown here can be used as an upper bound on the performance of this receiver.

The algorithm called decorrelating decision-feedback multiuser detector [23] has been modified and simulated on our CDMA system. For this algorithm, the order of detection is important. Thus, we ordered the users in an order of decreasing SNR at the decorrelating receiver output. The performance was inferior to the MMSE receiver for both channels A and B, and indeed if we look at the values of $F_{k,k}$ of (9) in that reference we observe 10–20-dB degradation for the last users. Since it is more complex than the MMSE, we conclude that this algorithm is not useful for the problem of this paper.

Finally, we like to add the simulation results for coded OFDM on channels A and B. The code used was $K = 7$ rate 1/2 convolutional code with random interleaver and optimal soft input, and the modulation was QPSK, see Fig. 11. The cyclic prefix

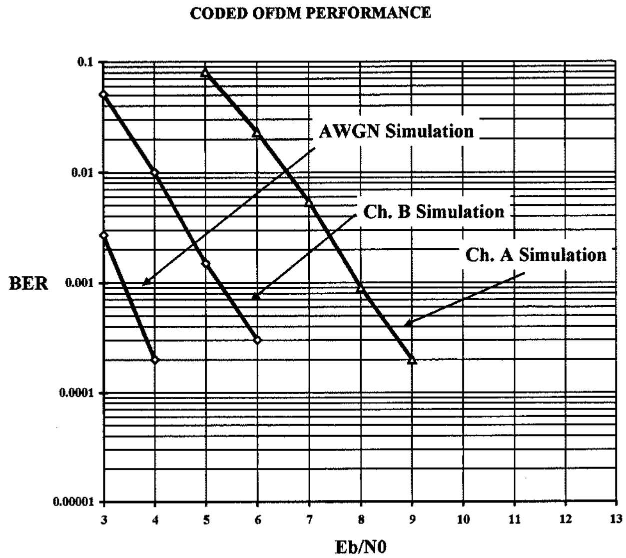



Fig. 11. Simulation results for coded OFDM with 64 states rate 1/2 code, random interleaver, known channel, coherent QPSK.

was of length 16, so larger than the channel length. The results are as follows. Channel A causes about 5-dB degradation relative to the same code on AWGN, while channel B causes about 2-dB degradation. Comparing this result to Fig. 6, we see that OFDM is slightly better on channel A but slightly worse on channel B. With MMSE initialization (not shown), the CDMA closes the gap on channel A and improves on channel B. A more rigorous comparison is needed in order to obtain a good conclusion.

## VI. Conclusion

We have presented an efficient multiuser detection method for CDMA on channels with multipath and white noise. The method is useful for handling channels which are distorting enough to cause most previous method to fail or suffer high degradation. We compared the nonlinear method with the optimal linear method. Comparison with several other known techniques are also provided. The results are very promising. They show that even with a large delay spread and a maximum number of users, there can be almost no degradation from the channel distortion for the large majority of randomly selected channels. We show that a small (several decibels) fading margin is sufficient for 1% outage probability provided that the multipath is already long enough to provide antifading diversity.

## Appendix A
## Linear MMSE Solution

The MMSE solution to the general problem of (5) is well known (e.g., [16, Ch. 15]). The received vector $\mathbf{r}$ is to be multiplied with the matrix

$$\begin{aligned} D &= (S^\dagger S + \sigma^2 I_K)^{-1} S^\dagger \\ &= (C^\dagger H^\dagger H C + \sigma^2 I_K)^{-1} C^\dagger H^\dagger. \end{aligned} \tag{29}$$

The *decorrelating receiver* is obtained by setting $\sigma = 0$ in (29), and is the least squares solution. In the trivial case of $K = 1$, we obtain the matched filter or RAKE receiver. In the case of $K = N$, $C$ is orthonormal and therefore

$$\begin{aligned} D &= (C^\dagger (H^\dagger H + \sigma^2) C)^{-1} C^\dagger H^\dagger \\ &= C^\dagger (H^\dagger H + \sigma^2)^{-1} H^\dagger. \end{aligned} \tag{30}$$

The meaning of the last result is that in this special case the problem is equivalent to an MMSE equalizer followed by a conventional despreader! Unfortunately for $1 < K < N$, no such simplification is possible. The reason that $K = N$ is a special case is that the vector $C\mathbf{u}$ becomes uncorrelated, hence the MMSE receiver for $C\mathbf{u}$ can be obtained independently of the specific value of $C$. In the case of the decorrelator, the case $K = N$ is equivalent to a zero-forcing equalizer in front of a conventional despreader.

## Appendix B
## Residual Error from ISI

Let us assume random codes with independent and identically distributed chips of variance $1/N$ and let user $k$ be in error at the previous symbol. The spillover signal resulting from this error is

$$\tilde{y}_i = e_k \sum_{j=i+1}^{L} h_j c_{i+N-j,\,k}, \qquad i = 0, \cdots, L-1 \tag{31}$$

where $e_k$ is the symbol error and takes the values $\pm 2$ or $\pm 2j$. The average total energy in this interference signal is

$$\begin{aligned} I &= E\left[\sum_{i=0}^{L-1} |\tilde{y}_i|^2\right] \\ &= 4 \sum_{i=0}^{L-1} E\left[\left|\sum_{j=i+1}^{L} h_j c_{i+N-j,\,k}\right|^2\right] \\ &= \frac{4}{N} \sum_{i=0}^{L-1} \sum_{j=i+1}^{L} |h_j|^2 \\ &= \frac{4}{N} \sum_{j=1}^{L} j |h_j|^2. \end{aligned} \tag{32}$$

Let

$$L_e = \frac{\sum_{j=1}^{L} j |h_j|^2}{\sum_{j=0}^{L} |h_j|^2}. \tag{33}$$

It is clear that $L_e \le L$, and for constant profile $L_e$ will be of the order of $L/2$. The ratio between a bit energy and the interference energy density due to $b$ detection errors (out of $2K$ bits of previous symbol) is

$$\frac{E_b}{I} = \frac{N^2}{4bL_e}. \tag{34}$$

In the case that decision feedback is not used, the resulting expression will be

$$\frac{E_b}{I} = \frac{N^2}{2KL_e}. \tag{35}$$

## REFERENCES


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

[2] Y. C. Yoon, R. Kohno, and H. Imai, "A spread-spectrum multiaccess system with cochannel interference cancellation," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 1067–1075, Sept. 1993.

[3] M. K. Varanasi and S. Vasudevan, "Multiuser detectors for synchronous CDMA communication over nonselective Rician fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 711–722, Feb./Mar./Apr. 1994.

[4] M. K. Varanasi and B. Aazhang, "Near optimum detection in synchronous code-division multiple-access systems," *IEEE Trans. Commun.*, vol. 39, pp. 725–736, May 1991.

[5] D. Divsalar, M. K. Simon, and D. Raphaeli, "A new approach to parallel interference cancellation for CDMA," in *Proc. GLOBECOM'96*, London, U.K., Nov. 1996.

[6] A.-L. Johansson and A. Svensson, "Multi-stage interference cancellation in multi-rate DS/CDMA systems," in *Proc. PIMRC'95*, Toronto, ON, Canada, Sept. 1995, pp. 965–969.

[7] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*, vol. 38, pp. 509–519, Apr. 1990.

[8] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 796–807, June 1994.

[9] S. L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, Feb./Mar./Apr. 1995.

[10] Z. Xie, R. Short, and C. Rushford, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.

[11] Z. Zvonar and D. Brady, "Suboptimal multiuser detector for frequency-selective rayleigh fading synchronous CDMA channels," *IEEE Trans. Commun.*, vol. 43, pp. 154–157, Feb./Mar./Apr. 1995.

[12] E. Colzi, C. Elia, F. Tarkoy, and R. Viola, "Synchronous CDMA for satellite frequency-selective broadcasting channels: Performance and receiver structures," in *Proc. ICC'93*, Geneva, Switzerland, May 1993, pp. 1746–1751.

[13] A. Franchi, C. Elia, and E. Colzi, "Maximum likelihood multipath channel estimation for synchronous-CDM systems," in *GLOBECOM'94*, San Francisco, CA, Nov. 28–Dec. 2, 1994, pp. 88–92.

[14] D. Raphaeli, "Nonlinear multiuser detector for synchronous coded CDMA on severely distorted frequency selective channels," in *Proc. GLOBECOM'96*, London, U.K., Nov. 1996.

[15] L. B. Nelson and H. V. Poor, "Iterative multiuser receivers for CDMA channels: An EM-based approach," *IEEE Trans. Commun.*, vol. 44, pp. 1700–1710, Dec. 1996.

[16] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[17] D. Raphaeli, "Iterative cochannel interference cancellation in synchronous CDMA on a frequency selective channel," in *Proc. ICUPC'96*, Cambridge, MA, Sept. 29–Oct. 2, 1996.

[18] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech Signal Processing*, vol. 36, pp. 477–489, Apr. 1988.

[19] S. Laufer and R. Hasid, Performance enhancement of CDMA networks, submitted for publication.

[20] S. H. Jamali and T. Le-Ngoc, *Coded Modulation Techniques for Fading Channels*. Boston, MA: Kluwer, 1994.

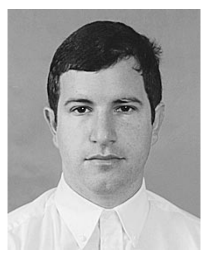
[21] D. Raphaeli, "The performance of synchronous CDMA on frequency selective channels," IEEE Trans. Commun., submitted for publication.

[22] N. Yee and J. Linnartz, "Controlled equalization of multi-carrier CDMA in an indoor Rician fading channel," in *Proc. IEEE Vehicular Technology Conf.*, 1994, pp. 1665–1669.

[23] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Trans. Commun.*, vol. 41, pp. 285–290, Feb. 1993.

[24] K. Rose and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1249–1257, July 1992.

[25] T. Graepel, M. Burger, and K. Obermayer, "Deterministic annealing for topographic vector quantization and self-organizing maps," in *Proc. WSOM'97*, Helsinki, Finland, June 1997.

**Dan Raphaeli** (S'93–M'95–SM'99) was born in Israel in 1967. He received the B.Sc. degree in electrical and computer engineering from Ben Gurion University, Israel, in 1986, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 1992 and 1994, respectively.

From 1986 to 1991, he was a Research Member at the Electronic Research Institute of the Israel Defence Ministry, where he was involved in the development of many advanced communication and signal processing projects. From 1992 to 1994, he was at the Jet Propulsion Laboratory, Pasadena, CA, where he was involved in research on communication systems for future spacecrafts. Since 1994, he has been an Assistant Professor at the Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv, Israel. He is also a Professional Consultant to industry in modem design for wireless and twisted-pair high-speed communications and DSP. His research subjects include modulation/demodulation, turbo codes, coding and decoding algorithms, spread spectrum, mobile communication, synchronization, equalization, and digital signal processing.