Improved Parallel Interference Cancellation for CDMA

Dariush Divsalar, *Fellow, IEEE*, Marvin K. Simon, *Fellow, IEEE*, and Dan Raphaeli, *Member, IEEE*

***Abstract*—This paper introduces an improved nonlinear parallel interference cancellation scheme for code-division multiple access (CDMA) that significantly reduces the degrading effect on the desired user of interference from the other users that share the channel. The implementation complexity of the scheme is linear in the number of users and operates on the fact that parallel processing simultaneously removes from each user a part of the interference produced by the remaining users accessing the channel the amount being proportional to their reliability. The parallel processing can be done in multiple stages. The proposed scheme uses tentative decision devices at the multiple stages to produce the most reliably estimated received data for generation and cancellation of user interference. Simulation results are given for a multitude of different situations, in particular, those cases for which the analysis is too complex.**

***Index Terms*—Communication theory, spread spectrum communicatons.**

I. Introduction

MULTIUSER communications systems that employ code-division multiple access (CDMA) exhibit a user capacity limit in the sense that there exists a maximum number of users that can simultaneously communicate over the channel for a specified level of performance per user. This limitation is brought about by the ultimate domination of the other user interference over the additive thermal noise. Over the years researchers have sought ways to extend the user capacity of CDMA systems either by employing optimum [maximum-likelihood (ML)] detection, interference cancellation (IC) methods, or other methods such as the decorrelating receiver [1]–[14].

With regard to the former, the work of Verdu [1], [2] is perhaps the most cited in the literature and the one upon which much of the other work is based. In Verdu's work, the receiver structure is derived based on minimizing the squared Euclidean distance between the received signal and the *sum* of the M asynchronous user signals, i.e., the total transmitted signal. As such, the presence of all M users simultaneously sharing the channel is accounted for in arriving at the ML receiver. The primary difference between the structure that evolves from such an approach and the conventional structure is that *joint sequence* decisions are made on the set of M matched filter outputs as opposed to individual bit-by-bit decisions on each matched filter output alone.

While indeed such optimum multiuser algorithms offer significantly improved performance by alleviating the disadvantages associated with the conventional scheme, they unfortunately suffer from the fact that their complexity grows exponentially with the number of users and the length of the sequence. This follows directly from the fact that the optimum ML decision algorithm can be implemented as a dynamic program with time complexity per binary decision that is $O(2^M)$ [15]. While in many practical applications such performance complexity prohibits implementation of the Verdu algorithm, its performance is still very much of interest since it serves as a benchmark against which to compare other schemes with less implementation complexity such as those that employ interference cancellation to be discussed shortly.

One disadvantage of most multiuser detectors including the proposed one is the necessity of knowing the relative amplitudes of the various user signals present at the input to the receiver. One possibility around this disadvantage is to perform multiuser amplitude estimation [16], or employ the decorrelator receiver which is not sensitive to the relative powers of the users. An alternative scheme is to employ power control at the transmitter which is a common technique used in cellular radio systems to solve the near–far problem. In this case, all received users are assumed to have the same power. (Note that in practical applications perfect power control is hard to achieve.) The most obvious solution to the multiuser interference problem would be to design the user codes to have more stringent cross correlation properties since indeed if the signals were truly orthogonal this interference would not exist. Unfortunately, it is not theoretically possible that any set of codes will exhibit zero cross correlation in the asynchronous case. Moreover, the near–far problem mentioned above still exists even for well-designed almost-orthogonal codes. Thus, the multiuser interference problem must be dealt with and tackled from another viewpoint.

One approach is to employ a suitable linear transformation on the matched filter outputs. Belonging to this family is the

Paper approved by C. Robertson, the Editor for Spread Spectrum Systems of the IEEE Communications Society. Manuscript received August 21, 1996; revised July 7, 1997. This work was carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with the National Aeronautics and Space Administration, and was supported by the Director's Research and Discretionary Fund at JPL and Qualcomm, Inc. This paper was presented in part at the IEEE Communication Theory Workshop, Santa Cruz, CA, April 24–26, 1995.

D. Divsalar and M. K. Simon are with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA (e-mail: marvin.k.simon@jpl.nasa.gov; dariush@shannon.jpl.nasa.gov).

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

Publisher Item Identifier S 0090-6778(98)01928-X.

so-called decorrelating receiver [4]. In this method, the different users are made uncorrelated by a linear transformation. This linear transformation is computed by measuring all cross correlations between pairs of user codes and then inverting the resulting (typically huge) matrix of cross correlations. Since in practical systems each user is assigned a very long peudonoise (PN) code, each bit has essentially a random code assigned to it. Thus, in this case, the above procedure would have to be repeated for each bit in succession! In the asynchronous and/or time dispersed case the complexity of the computation is increased still further.

Another popular approach is to employ interference cancellation, i.e., to attempt removal of the multiuser interference from each user's received signal before making data decisions. In principle, the IC schemes considered in the literature fall into two categories, namely, *serial (successive)* and *parallel* cancellation. With regard to the former, Viterbi [6] (see also Dent [7] and Patel and Holtzman [19]) suggested coordinated processing of the received signal with a successive cancellation scheme in which the interference caused by the remaining users is removed from each user in succession. One disadvantage of this scheme is the fact that a specific geometric power distribution must be assigned to the users in order that each see the same signal power to background plus interference noise ratio. This comes about because of the fact that with successive cancellation the first user to be processed sees all of the interference from the remaining $M-1$ users, whereas each user downstream sees less and less interference as the cancellation progresses. Another disadvantage of this scheme has to do with the required delay necessary to fully accomplish the IC for all users in the system. Since the IC proceeds serially, a delay on the order of M computation stages is required to complete the job. Nevertheless, Viterbi showed that the successive IC scheme could approach channel capacity for the aggregate Gaussian noise channel. As such, the scheme does not become multiuser interference limited.

Parallel processing of multiuser interference *simultaneously* removes from *each* user the interference produced by the remaining users accessing the channel. In this way, each user in the system receives equal treatment insofar as the attempt is made to cancel his or her multiple-user interference. As compared with the serial processing scheme, since the IC is performed in parallel for all users, the delay required to complete the operation is at most a few bit times. The early papers that dealt with parallel IC recognized the desire to arrive at a structure that could be motivated by the ML approach. In particular, a multistage iterative approach was suggested by Varanasi and Aazhang [8], [9] which at a given stage estimated a given user's bit under the assumption that the exact knowledge of the other users' bits in the same transmission interval needed to compute the multiuser interference could be replaced by *estimates* of these bits from the previous stage. It was indeed this basic idea which led to the multistage iterative schemes subsequently proposed by Yoon, Kohno, and Imai [11]–[13] and Kawabe *et al.* [14]. What was common to all of these schemes was the fact that at each stage of the iteration, an attempt was made for each user to *completely* cancel the interference caused by all the other users.[1] As we shall see in this paper, this is not necessarily the best philosophy. Rather, when the interference estimate is poor (as in the early stages of interference cancellation), it is preferable not to cancel the entire amount of estimated multiuser interference.[2] As the IC operation progresses, the estimates of the multiuser interference improve and, thus, in the later stages of the iterative scheme, it becomes desirable to increase the weight of the interference being removed. The motivation behind this approach can also be derived from ML considerations as was done for the total IC approach previously considered.

With the above discussion in mind, this paper presents a new parallel interference cancellation scheme that significantly reduces the degrading effect of multiuser interference but with a complexity linear in the number of users and with improved performance over the previously considered parallel and serial processing techniques. When compared with classical CDMA without IC, the improvement in performance is dramatic at the expense of a practically feasible increase in complexity. Although our scheme (as well as the other schemes mentioned) is suitable to the case of a nonuniform power distribution as well as a uniform power distribution among the users, in this paper we shall primarily focus on the latter. In addition, although ours and the other parallel schemes are applicable to asynchronous transmission with no increase of complexity, we shall assume here that all users have synchronous data streams. This case results in worst case performance, i.e., if the data transition instants of the various users are not aligned, then on the average they have less of an interfering effect on one another. This happens since the subchip intervals produced by the multiplication of two sequences spreads into a wider spectrum and their integral is on the average lower. Note, that in the asynchronous time/nonuniform power case the estimation of more parameters is needed, namely, the delays of the users and their powers. This parameter estimation problem is common to most of the above techniques.

II. MULTIUSER COMMUNICATION SYSTEM MODEL

We consider a CDMA communication system in which M users are communicating simultaneously at the same rate over a common additive white Guassian noise (AWGN) channel each with a binary phase-shift keying (BPSK) data modulation and its own PN code. As such, the received signal is the sum of M direct sequence BPSK signals each with power S_i, bit time T_b, and PN chip time T_c, and AWGN with single-sided power spectral density (PSD) N_0 W/Hz. At baseband, this signal can be written in the complex form[3]

$$r(t) = \sum_{i=1}^{M} s_i(t) + n(t) = \sum_{i=1}^{M} \sqrt{S_i}\, m_i(t) PN_i(t) e^{j\phi_i} + n(t) \tag{1}$$

[1] We shall refer to such a technique as *brute force* or *total* interference cancellation.

[2] We shall refer to such a technique as *partial* interference cancellation.

[3] For convenience, we shall use complex notation to represent the various signals in the receiver.

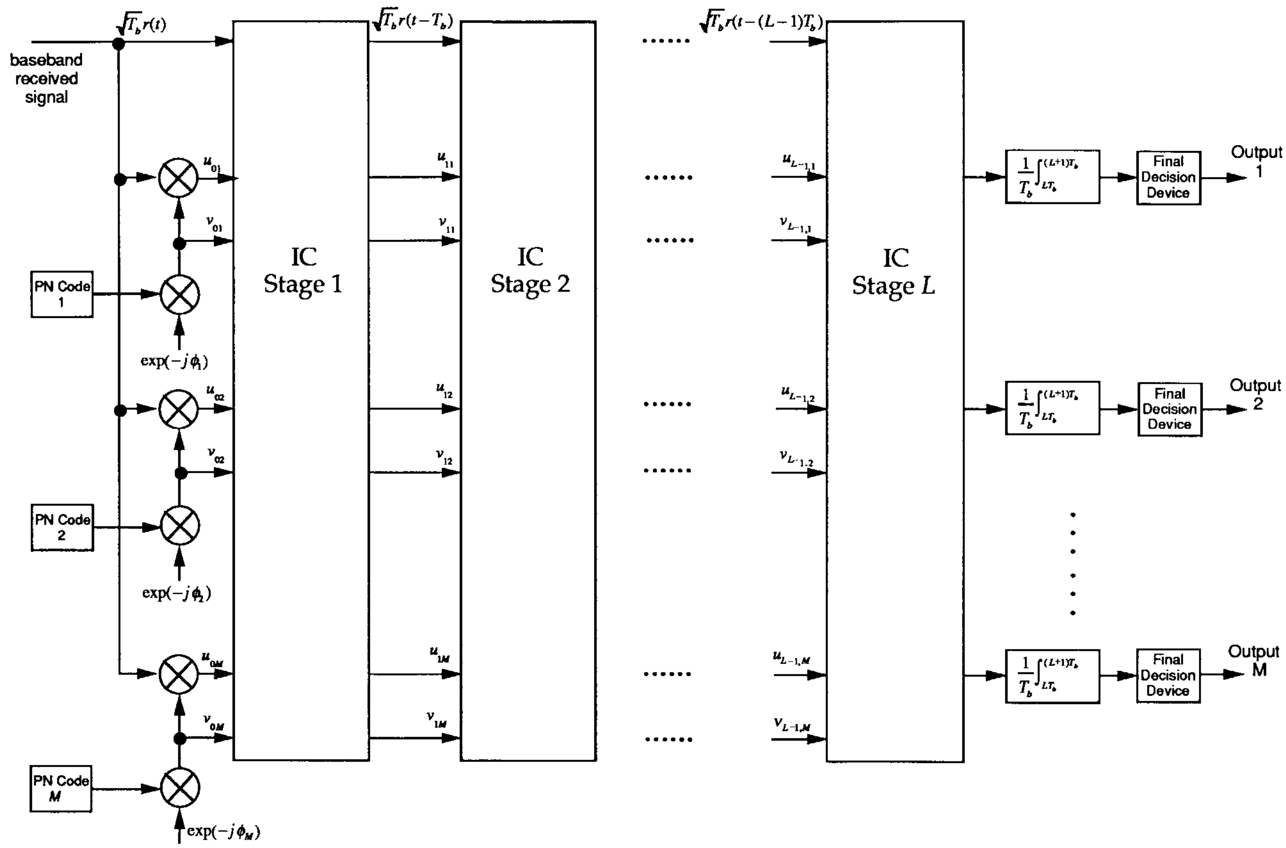

Fig. 1. An L-stage interference cancellation scheme with parallel processing for CDMA (complex baseband model).

where, for the ith user, $PN_i(t)$ is the PN code waveform, $m_i(t) = \sum_{k=-\infty}^{\infty} a_{ik} p(t - kT_b)$ is the data modulation with kth bit a_{ik} taking on equiprobable values ± 1 and unit power rectangular pulse shape $p(t)$ of duration T_b, and ϕ_i is the carrier phase. For the equal user power case, one would have $S_i = S$; $i = 1, 2, \cdots, M$. As previously mentioned, one must provide a means (not discussed in this paper) to estimate the user's power, the carrier phase, and, in the asynchronous case, the code delays.

We shall assume for the purpose of analysis and simulation that the users have purely random PN codes assigned to them. This assumption is justified by the use of long PN codes, with period much longer than the bit duration, as, for example, in the IS-95 cellular mobile standard. It is to be emphasized, however, that the IC schemes to be discussed in what follows apply equally well to any appropriate set of PN codes chosen for the users, provided that the codes are known to the receiver. In view of our assumption, over the bit interval $0 \le t \le T_b$, the ith user's PN waveform can be expressed in the form

$$PN_i(t) = \sum_{k=1}^{\eta} c_{ik} p(t - kT_c) \tag{2}$$

where $p(t)$ is again a unit power rectangular pulse shape now of duration T_c, $\eta = T_b/T_c$ is the number of PN code chips per data bit, i.e., the spreading ratio, and $\{c_{ik}\}$ is a random binary (± 1) sequence. For a given set of user codes, their normalized cross correlation matrix $\mathbf{\Gamma} = [\gamma_{ij}]$ is defined as

$$\gamma_{ij} = \frac{1}{\eta} \sum_{k=1}^{\eta} c_{ik} c_{jk} = \frac{1}{T_b} \int_0^{T_b} PN_i(t) PN_j(t)\, dt, \qquad i, j = 1, 2, \cdots, M \tag{3}$$

with $\gamma_{ii} = 1$; $i = 1, 2, \cdots, M$. Note that computation of the cross correlation matrix is not needed for the receiver developed in this paper.

III. Derivation of the New Parallel Iterated Multiuser Detector

As previously stated, the optimal multiuser detector [1] is derived from a joint ML decision on the M user data bits in a given interval and thus has exponential complexity in the number of users. Instead, we choose to *individually* decide on each user's data bit in this same interval, the motivation being to reduce the complexity of the detector. Clearly, in deriving such an ML metric for any one user, one would theoretically need exact knowledge of the data bits corresponding to all of the other $M - 1$ users. Since indeed this information is unknown, the above *theoretical* assumption is *practically* invalid. However, by replacing exact knowledge of the other $M - 1$ user bits by estimates of their values, we arrive at an iterative scheme (see Fig. 1) wherein each stage of the iteration produces new and better estimates of the user bits

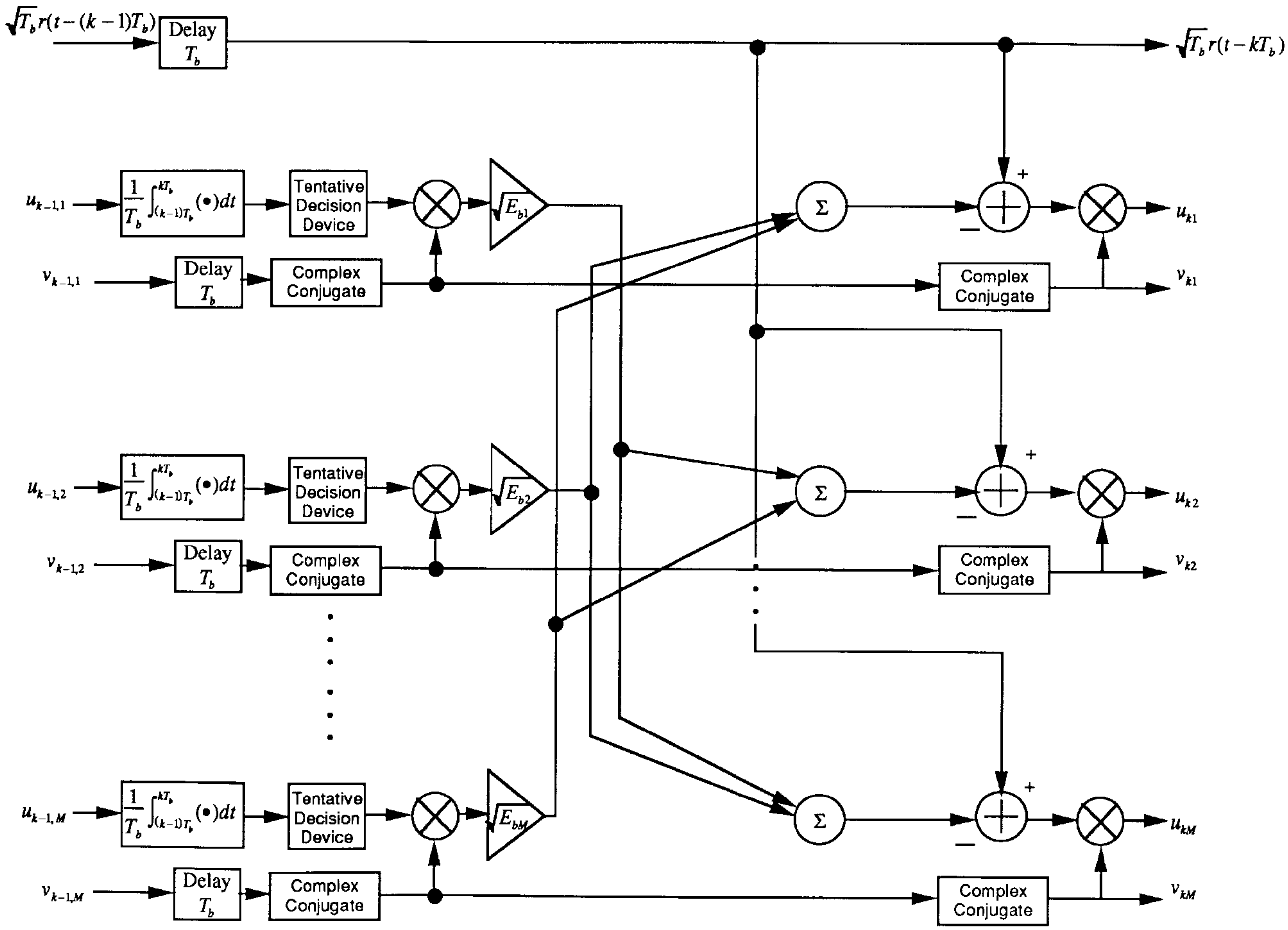

Fig. 2. kth stage of total interference cancellation.

based upon those obtained in the previous stage. By solving this ML problem with the above condition imposed, we arrive at the following decision rule for user 1's (assumed herein to be the user of interest) bit at the kth iteration stage[4] (see Fig. 2)

$$\begin{aligned}\hat{a}_1(k) &= \operatorname{sgn}\Bigg\{\operatorname{Re}\Bigg[\mathrm{e}^{-\mathrm{j}\phi_1}\mathrm{y}_1 \\ &\quad - \sum_{i=2}^{M}\sqrt{\frac{2E_{bi}}{N_0}}\,\hat{a}_i(k-1)e^{j(\phi_i-\phi_1)}\gamma_{1i}\Bigg]\Bigg\} \\ &\triangleq \operatorname{sgn}\{\mathrm{Y}_1 - \hat{\mathrm{I}}_1(\mathrm{k})\} \\ \hat{a}_1(0) &= \operatorname{sgn}\{\mathrm{Y}_1\}\end{aligned} \tag{4}$$

where $Y_1 = \operatorname{Re}\{\mathrm{e}^{-\mathrm{j}\phi_1}\mathrm{y}_1\}$ with

$$y_i = \sqrt{\frac{2}{N_0 T_b}}\int_0^{T_b} r(t)PN_i(t)\,dt \tag{5}$$

i.e., a normalized projection of the received signal on user i's code, $E_{bi} = S_i T_b$ is the bit energy in user i's signal, $\{\hat{a}_i(k-1)\}|_{i=1}^{M}$ are the estimated data bits from the previous iteration, and $\hat{I}_1(k)$ denotes the estimated interference contributed by the other users to user 1. The decision represented by (4) is referred to as a *tentative decision* (for any stage previous to the last) since indeed the final decision on user 1's data bit is only made at the last iteration stage, i.e., after the interference has been removed to whatever extent is possible.

In the above scheme, insofar as detection of user 1's data bit is concerned (or, for that matter, any other user of interest's data bit) the *total* interference is estimated and canceled from the received signal at each stage of the iteration. Since in the early stages of interference cancellation the tentative decisions are less reliable than they are in later stages, it is not intuitively clear that the above philosophy of entirely canceling the full amount of interference at each iteration stage necessarily leads to the best suboptimum decision metric. Rather, a better philosophy is one which in the early stages cancel only a fraction of the multiuser interference with the amount being canceled increasing as one continues to iterate toward the ultimate final data decisions, i.e., as the fidelity of the tentative decisions improves. Viewing the receiver as an iterative algorithm, what we do, in effect, is control the step size of the algorithm. To see how a metric motivated by such a philosophy can come about we proceed as follows.

We first note that in arriving at the estimate of a_1 (the user of interest's bit) in (4), the information on this bit (e.g., its

[4]The details of the derivation leading to (4) are given in [18]. For simplicity of notation, we shall drop the second subscript on user 1's data bit and assume that we are considering only the interval $0 \le t \le T_b$, i.e., we shall denote a_{10} by a_1 and likewise for the decisions on this bit.

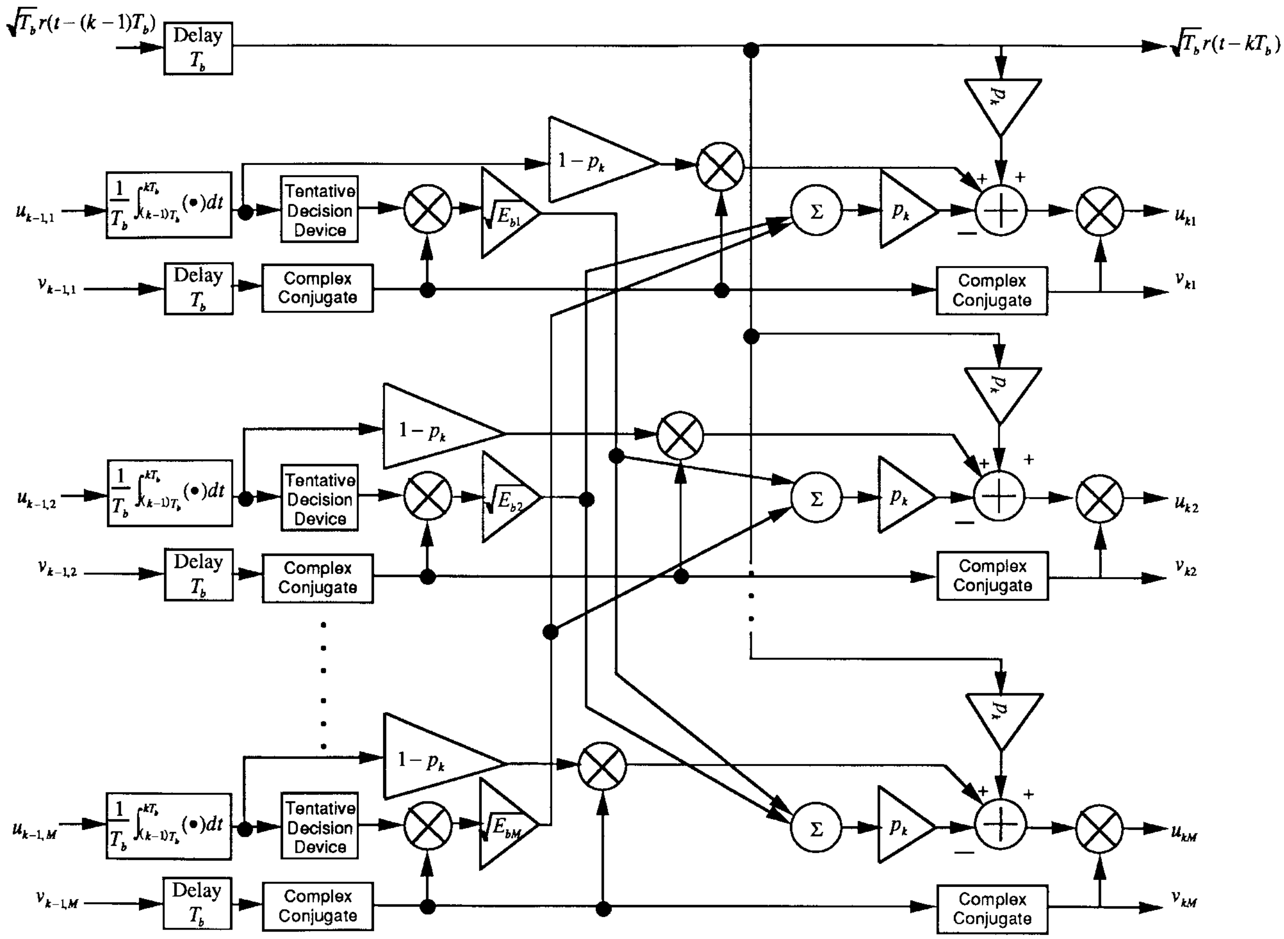

Fig. 3. kth stage of partial interference cancellation.

estimate) available at the previous iteration stage has not been used at all. By including this information from the previous iteration, namely, the tentative hard decision $\hat{a}_1(k-1)$, or, better yet, the soft (linear) tentative decision $\tilde{a}_1(k-1)$ (i.e., the value just before the decision device), one can improve upon the above IC technique. In particular, the improved soft (linear) decision $\tilde{a}_1(k)$ at the kth iteration stage will be obtained as a *weighted sum* of $\tilde{a}_1(k-1)$ and $[Y_1 - \hat{I}_1(k-1)]$. The tentative bit estimate $\hat{a}_1(k)$ is then obtained by passing $\tilde{a}_1(k)$ through a tentative decision device. Interestingly enough, the suggestion to form such a weighted sum comes from considerations based on *jointly* observing Y_1 and $\tilde{a}_1(k-1)$ as opposed to Y_1 alone, and leads to an iterative structure analogous to Fig. 1 with kth stage as in Fig. 3 and conceptual equivalent as in Fig. 4. In particular, consider the component of the normalized received signal vector corresponding to user 1, i.e., y_1 of (5), which, when multiplied by $e^{-j\phi_1}$, is obtained from (1)–(5) as

$$e^{-j\phi_1}y_1 = \sqrt{\frac{2E_{b1}}{N_0}}\, a_1 + \sum_{i=2}^{M} \sqrt{\frac{2E_{bi}}{N_0}}\, a_i e^{j(\phi_i-\phi_1)}\gamma_{1i} + n_1 e^{-j\phi_1} \tag{6}$$

where n_1 is a zero-mean variance two (unit variance per dimension) complex Gaussian random variable (RV). Taking the real part of (6) gives an expression of the form

$$Y_1 = \sqrt{\frac{2E_{b1}}{N_0}}\, a_1 + I_1 + N_1 \triangleq \sqrt{\frac{2E_{b1}}{N_0}}\, a_1 + \hat{I}_1 + W_1; \quad W_1 = I_1 - \hat{I}_1 + N_1 \tag{7}$$

where N_1 is a zero-mean unit variance Gaussian RV, I_1 denotes the real part of the multiuser interference experienced by user 1 due to the remaining $M-1$ users, and $\hat{I}_1$ denotes an estimate of I_1 based on estimates of the other user data bits. As such, $I_1 - \hat{I}_1$ represents the *residual* (uncanceled) real multiuser interference. Since the estimates of the other user data bits are not available at the time that user 1's data bit is being estimated, (6) and (7) suggest, as previously discussed, an iterative (multistage) structure in which the other user data bit estimates are obtained from the previous stage. As such, (7) can be rewritten as

$$Y_1 = \sqrt{\frac{2E_{b1}}{N_0}}\, a_1 + \hat{I}_1(k) + W_1(k) \tag{8}$$

where

$$W_1(k) = I_1 - \hat{I}_1(k) + N_1; \quad \hat{I}_1(k) \triangleq \mathrm{Re}\left\{\sum_{i=2}^{M} \sqrt{\frac{2E_{bi}}{N_0}}\, \hat{a}_i(k-1) e^{j(\phi_i-\phi_1)}\gamma_{1i}\right\} \tag{9}$$

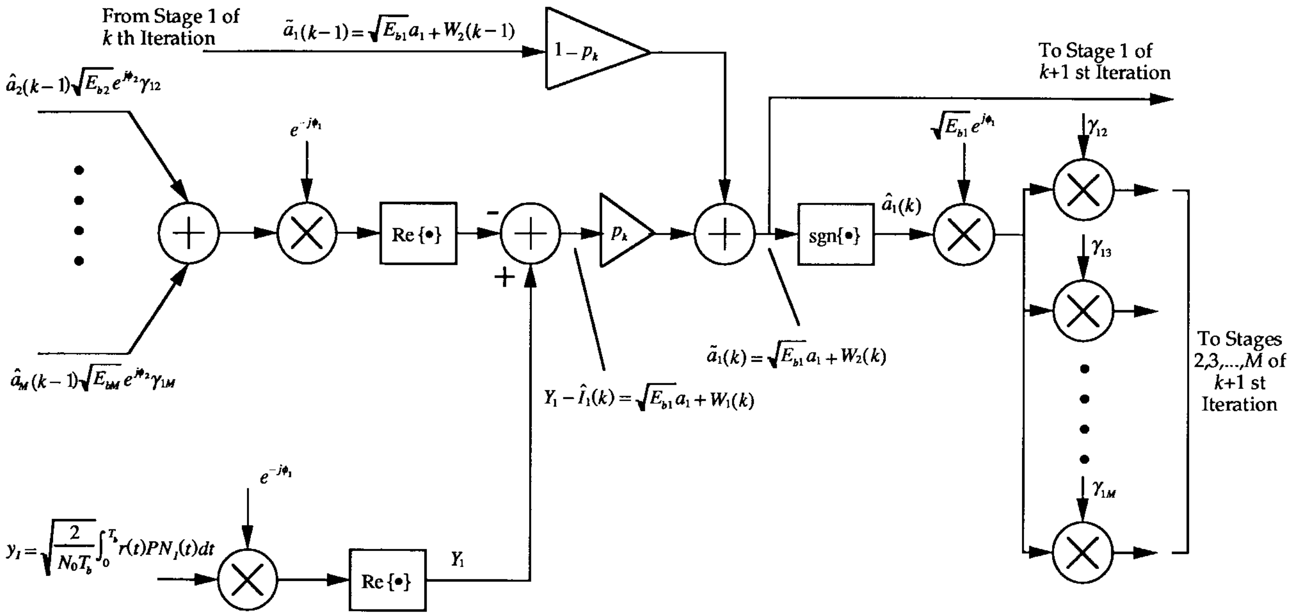

Fig. 4. Conceptual block diagram of stage 1 of kth iteration.

with the notation "(k)" referring as in the previous discussions to the kth stage. For the purpose of what follows we can model the residual interference term of (9) as a zero-mean Gaussian RV which, when combined with N_1, results in a zero-mean Gaussian RV $W_1(k)$ with variance $E\{W_1^2(k)\} = \sigma_{1k}^2$. Note that the variance σ_{1k}^2 depends on the iteration stage k.

We recursively define $\tilde{a}_1(k)$ as the log-likelihood ratio (LLR) for a_1 given $\tilde{a}_1(k-1)$, Y_1, and $\hat{I}(k)$. It will be convenient to normalize this LLR such that its conditional average is equal to $\sqrt{2E_{b1}/N_0}\,a_1$. Note that under the Gaussian assumptions to follow, this LLR is also identical (up to a constant factor) to the *linear* minimum mean-square error estimate of $\sqrt{2E_{b1}/N_0}\,a_1$. For the first iteration no additional information is available; thus, we set $\tilde{a}_1(0) = Y_1$. We call $\tilde{a}_1(k)$ the *tentative* soft decision for user 1 generated at the kth iteration stage. As we will show in the following, a decision or nonlinear estimation of a_1 at each stage (using the Gaussian model) is simply a function of $\tilde{a}_1(k)$. This function is the tentative decision device mentioned earlier. For the sake of the derivations to follow, we shall assume that it is possible to approximate $\tilde{a}_1(k-1)$ of the previous stage as a conditional (on the data bit a_1) Gaussian RV, namely

$$\tilde{a}_1(k-1) = \sqrt{\frac{2E_{b1}}{N_0}}\,a_1 + W_2(k-1) \tag{10}$$

where $E\{W_2(k-1)\} = 0$, $E\{W_2^2(k-1)\} = \sigma_{2k}^2$. The Gaussian RV $W_2(k-1)$ depends on thermal noise N_1 and also on residual interferences from previous iterations. Our intent is to obtain an iterative solution for $\tilde{a}_1(k)$ given Y_1, $\tilde{a}_1(k-1)$ and $\hat{I}_1(k)$.

At a minimum, because both $W_2(k-1)$ and $W_1(k)$ contain the same thermal noise component, they are correlated, i.e., $E\{W_1(k)W_2(k-1)\} \triangleq \rho_k\sigma_{1k}\sigma_{2,k-1}$. As we shall see shortly, it is not necessary to be able to specifically evaluate σ_{1k}, $\sigma_{2,k-1}$, and ρ_k. Rather, a specific combination of these parameters will be used to define a parameter p_k which shall have significance in terms of the amount of interference for which cancellation is attempted at each stage of the receiver.

Using (8) and (10), the joint conditional probability density function (pdf) of Y_1 and $\tilde{a}_1(k-1)$ is given by (11), shown at the bottom of the page, which upon simplification becomes

$$\begin{aligned} p[Y_1,\, \tilde{a}_1(k-1)|a_1,\, \hat{I}_1(k)] &= \frac{1}{2\pi\sigma_{1k}\sigma_{2,k-1}\sqrt{1-\rho_k^2}} \\ &\cdot \exp\left(-\left\{\frac{\sigma_{2,k-1}^2\left[Y_1 - \sqrt{\frac{2E_b}{N_0}}\,a_1 - \hat{I}_1(k)\right]^2}{2\sigma_{1k}^2\sigma_{2,k-1}^2(1-\rho_k^2)}\right.\right. \\ &+ \left.\left.\frac{\sigma_{1k}^2\left[\tilde{a}_1(k-1) - \sqrt{\frac{2E_b}{N_0}}\,a_1\right]^2 - 2\rho_k\sigma_{1k}\sigma_{2,k-1}\left[Y_1 - \sqrt{\frac{2E_b}{N_0}}\,a_1 - \hat{I}_1(k)\right]\left[\tilde{a}_1(k-1) - \sqrt{\frac{2E_b}{N_0}}\,a_1\right]}{2\sigma_{1k}^2\sigma_{2,k-1}^2(1-\rho_k^2)}\right\}\right) \end{aligned} \tag{11}$$

(12), shown at the bottom of the page, where the constant C includes terms that do not depend on a_1. We now introduce a normalization so as to make the coefficients of $Y_1 - \hat{I}_1(k)$ and $\tilde{a}_1(k-1)$ in (12) sum to unity. Letting[5]

$$\begin{aligned} p_k &\triangleq \frac{\sigma_{2,k-1}^2 - \rho_k\sigma_{1k}\sigma_{2,k-1}}{\sigma_{1k}^2 + \sigma_{2,k-1}^2 - 2\rho_k\sigma_{1k}\sigma_{2,k-1}} \\ 1 - p_k &= \frac{\sigma_{1k}^2 - \rho_k\sigma_{1k}\sigma_{2,k-1}}{\sigma_{1k}^2 + \sigma_{2,k-1}^2 - 2\rho_k\sigma_{1k}\sigma_{2,k-1}} \end{aligned} \tag{13}$$

then (12) can be written in the desired form

$$\begin{aligned} &p[Y_1, \tilde{a}_1(k-1)|a_1, \hat{I}_1(k)] \\ &\quad = C\exp\left\{a_1\sqrt{\frac{2E_{b1}}{N_0}}\left[\frac{\sigma_{1k}^2 + \sigma_{2,k-1}^2 - 2\rho_k\sigma_{1k}\sigma_{2,k-1}}{\sigma_{1k}^2\sigma_{2,k-1}^2(1-\rho_k^2)}\right]\right. \\ &\qquad \left.\cdot\{p_k[Y_1 - \hat{I}_1(k)] + (1-p_k)\tilde{a}_1(k-1)\}\right\} \\ &\quad = C\exp(a_1\alpha_k\lambda_k) \end{aligned} \tag{14}$$

where

$$\begin{aligned} \alpha_k &\triangleq \sqrt{\frac{2E_{b1}}{N_0}}\left[\frac{\sigma_{1k}^2 + \sigma_{2,k-1}^2 - 2\rho_k\sigma_{1k}\sigma_{2,k-1}}{\sigma_{1k}^2\sigma_{2,k-1}^2(1-\rho_k^2)}\right] \\ \lambda_k &\triangleq p_k[Y_1 - \hat{I}_1(k)] + (1-p_k)\tilde{a}_1(k-1). \end{aligned} \tag{15}$$

Consider the LLR for bit a_1 as

$$\begin{aligned} L_1(k) &= \ln\frac{p[a_1 = 1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)]}{p[a_1 = -1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)]} \\ &= \ln\frac{\Lambda(1)p(a_1=1)}{\Lambda(-1)p(a_1=-1)} \\ &= \ln\frac{\Lambda(1)}{\Lambda(-1)} = 2\alpha_k\lambda_k. \end{aligned} \tag{16}$$

The mean of $\tilde{a}_1(k-1)$ is obtained from (10) as $\sqrt{2E_{b1}/N_0}\,a_1$ and the mean of $L_1(k)$ given a_1 is obtained from (16) as $2\alpha_k\sqrt{2E_{b1}/N_0}\,a_1$. Since from (10) the mean of $\tilde{a}_1(k)$ is $\sqrt{2E_{b1}/N_0}\,a_1$, then dividing $L_1(k)$ by $2\alpha_k$ gives the relation

$$\tilde{a}_1(k) = \lambda_k = p_k[Y_1 - \hat{I}_1(k)] + (1-p_k)\tilde{a}_1(k-1). \tag{17}$$

In addition to the linear estimation above, we would like to obtain a *nonlinear* estimate of a_1 for the purpose of obtaining the best reproduction of the other users' interference to be used for subtraction. An estimation of a_1 is possible using the ML principle, or using minimum mean square estimation (MMSE). Since the natural logarithm is a monotonic function of its argument, then taking the natural logarithm of (14), we see that the ML value of a_1 at the kth stage of iteration is simply given by

$$\begin{aligned} \hat{a}_1(k) &= \mathrm{sgn}\{\alpha_k\lambda_k\} \\ &= \mathrm{sgn}\{p_k[Y_1 - \hat{I}_1(k)] + (1-p_k)\tilde{a}_1(k-1)\} \\ &= \mathrm{sgn}\{\tilde{a}_1(k)\}. \end{aligned} \tag{18}$$

The ML approach produces a decision metric in which the tentative decisions at each stage of the iteration are hard decisions. The MMSE estimator (derived in the Appendix) is

$$\hat{a}_1(k) = \tanh(\alpha_k\lambda_k). \tag{19}$$

Comparing (19) with (18), we see that the hard tentative decisions have been replaced by soft tentative decisions in the form of hyperbolic tangent functions. Furthermore, the slope of these functions (which is proportional to α_k) is another parameter to be optimized at each stage of the iteration. For the simulations reported in this paper, we used only the hard decision approximation, although better performance can be obtained with the hyperbolic tangent function. At the last stage the best decision function for the uncoded case is the hard decision whereas, for the soft input coded case, using the LLR output $[\tilde{a}_1(k)]$ is recommended.

Comparing (18) with (4) we observe that the *weight factor* in the kth iteration p_k represents the amount of partial cancellation attempted at that stage and is a parameter to be optimized for each value of k. Intuitively, one would expect that the value of p_k (which depends on the particular stage through the subscript k) would monotonically increase as one progresses toward the final data decision, i.e., as one iterates more and more, the fidelity of the tentative decisions improves and, thus, one should attempt to weight more the interference cancellation term. Indeed, the numerical results to be presented later on bear out this intuition.

In summary, Fig. 1 is a multistage receiver structure with kth stage as in Fig. 2, which together result from the partial IC cancellation scheme suggested by (17) and (18) or (19). The modification of each iteration stage resulting from a brute force (total IC) approach such as that in Fig. 2, as suggested in [8]–[14] and (4) of this paper, is the inclusion of a parameter p_k to allow for *partial* cancellation of the multiuser interference at the kth stage and also the inclusion of estimated data for the user of interest obtained from the previous iteration. With regard to the form of the tentative decision device used at each iteration stage for each user, there are several options (see Fig. 5). A hard decision on user 1's data bit can, as

[5] Note that based on its definition, the parameter p_k is not necessarily restricted to lie in the range $0 \leq p_k \leq 1$. However, if $\rho_k < \min(\sigma_{1k}/\sigma_{2,k-1}, \sigma_{2,k-1}/\sigma_{1k})$ then it can be shown that this restriction is valid. This implies either $\sigma_{1k} \leq \sigma_{2,k-1}$ or $\sigma_{2,k-1} < \sigma_{1k}$. Intuition, however, would suggest the former.

$$\begin{aligned} &p[(Y_1, \tilde{a}_1(k-1)|a_1, \hat{I}_1(k)] \\ &\quad = C\exp\left(a_1\sqrt{\frac{2E_{b1}}{N_0}}\left\{\frac{(\sigma_{2,k-1}^2 - \rho_k\sigma_{1k}\sigma_{2,k-1})[Y_1 - \hat{I}_1(k)] + (\sigma_{1k}^2 - \rho_k\sigma_{1k}\sigma_{2,k-1})\tilde{a}_1(k-1)}{\sigma_{1k}^2\sigma_{2,k-1}^2(1-\rho_k^2)}\right\}\right) \end{aligned} \tag{12}$$

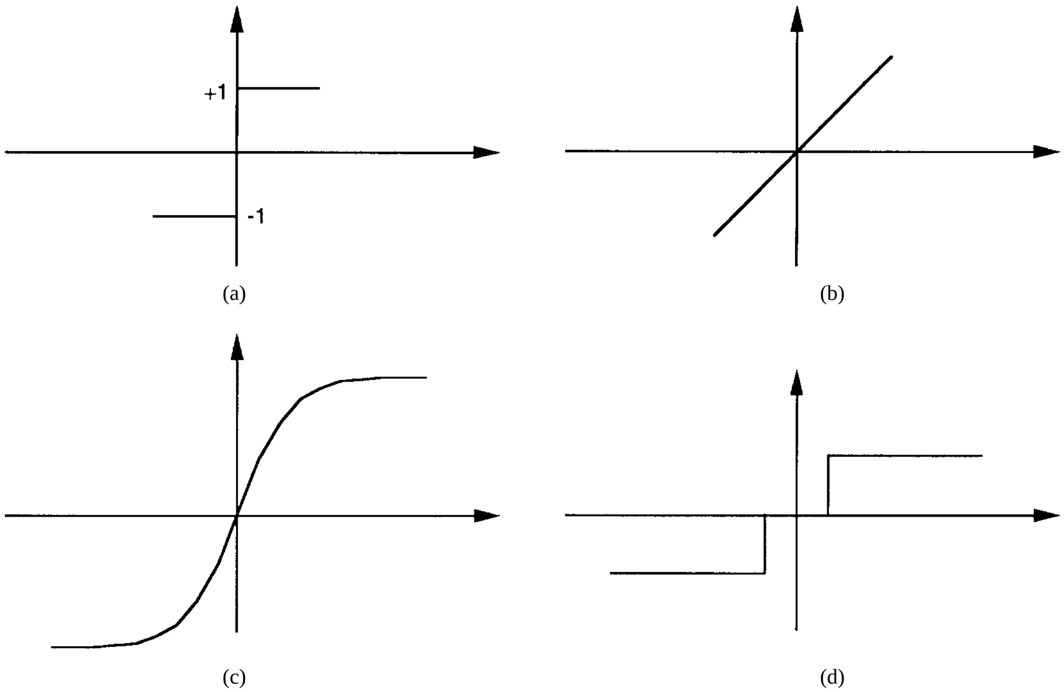

Fig. 5. Tentative decision devices. (a) Hard-limited (one-bit quantizer). (b) Linear (infinite-bit quantizer). (c) Hyperbolic tangent (soft quantizer). (d) Null zone device.

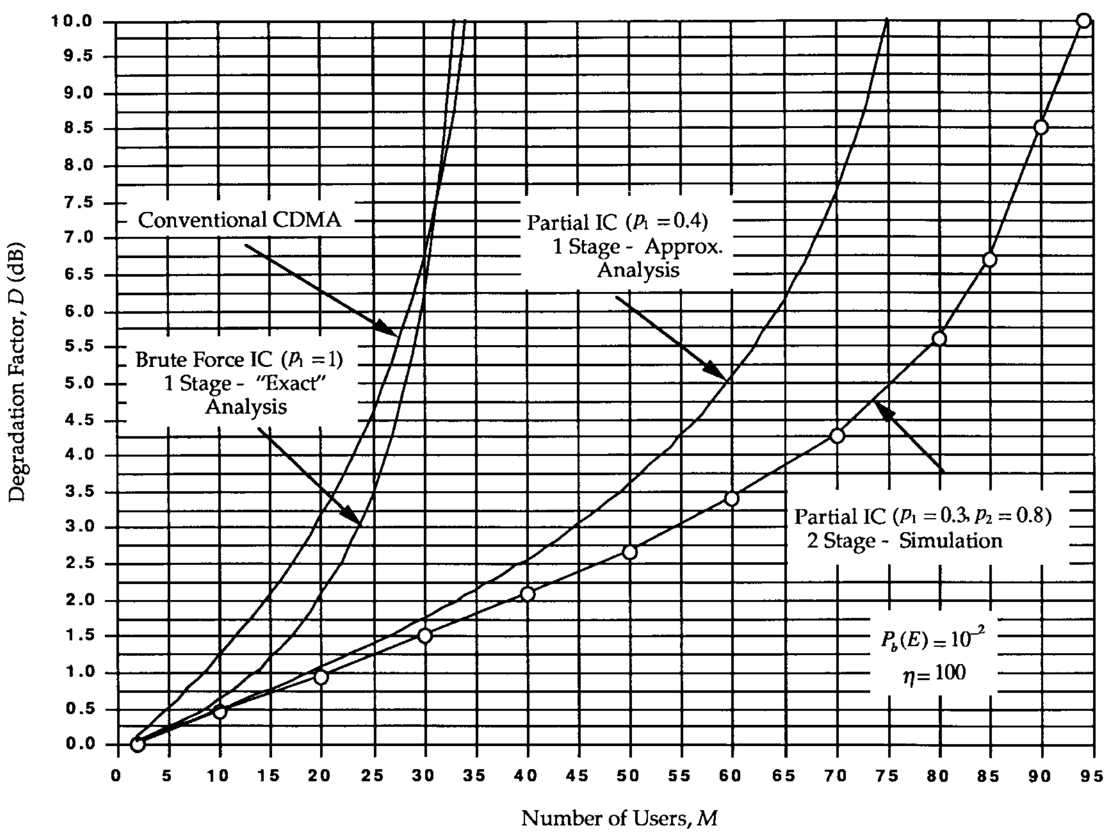

Fig. 6. A comparison of the degradation factors for one- and two-stage linear interference cancellation—equal power users.

described by (18), provide a tentative bit polarity for the next iteration. However, from a performance standpoint, it is better to provide soft tentative decisions to the succeeding iteration stages. Among the possibilities for softer decisions one can use are linear, null zone, and hyperbolic tangent devices. For example, the linear decision requires neither power estimates nor carrier demodulation; hence, a differential detection scheme can be employed instead of the coherent

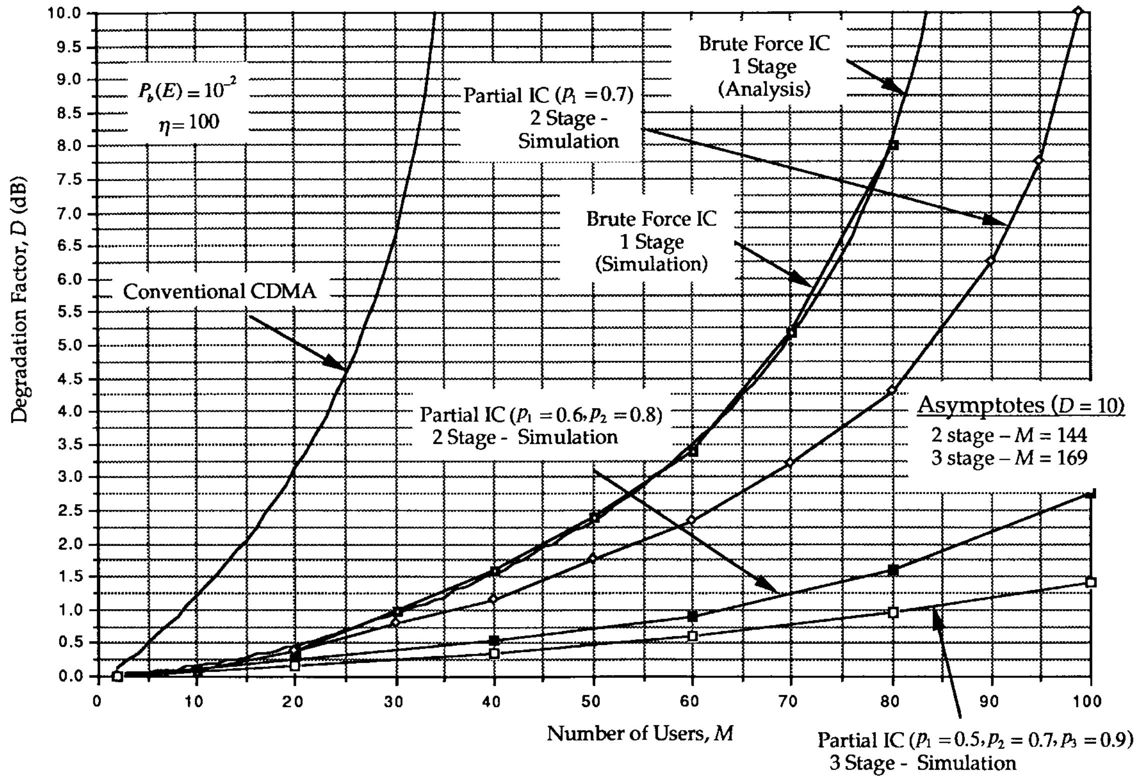

Fig. 7. A comparison of the degradation factors for one-, two-, and three-stage nonlinear interference cancellation—equal power users.

detection scheme assumed here. The null zone device provides slightly better performance than the hard decision device but is still inferior to the hyperbolic tangent device which is optimal from MMSE considerations based on a Gaussian interference assumption.

IV. Performance Results

The analysis of the performance of a single-stage interference cancellation is given in [18] and is quite tedious to obtain despite many simplifying but valid assumptions. Extending the analysis to only two stages is even more complicated; hence, it is expeditious to obtain the performance of a k-stage iterative partial IC scheme from computer simulations. Software programs have been written to model user transmitters and the base station receiver in the complex baseband domain. Random PN codes are generated for each user and used to spread his or her random data bits. The results of these spreading operations are multiplied by the complex form of the carrier phases $e^{j\phi_i}|_{i=1}^{M}$ following which complex Gaussian noise samples are added to the combined received signal with at least one sample per chip time. The carrier phases which are generated independently for each user are assumed to be constant over the integration time of the detector and uniformly distributed in the interval $(0, 2\pi)$.

It is common in analyses of CDMA systems [17] to define a *degradation factor* D as the ratio (in dB) of the E_b/N_0 required to achieve a given bit-error rate in the presence of M users, to that which would be required to achieve the same level of performance if only a single user was communicating. The performance results to be illustrated are plots of this degradation factor versus the number of users M for fixed values of processing gain η and bit-error probability $P_b(E)$.

Partial IC schemes were simulated with optimized values of the partial cancellation parameters p_k up to three stages.[6] The performance results described above are illustrated in Fig. 6 for linear and in Fig. 7 for nonlinear (hard decision) tentative decision devices along with the corresponding simulation and analysis results for a single-stage brute-force (total cancellation) scheme. Also illustrated are the results for conventional CDMA with no interference cancellation. We observe that, for the parameters considered and uncoded BPSK users, a three-stage nonlinear partial IC scheme allows as many as 80 users with a degradation of only 1 dB, as compared to nine users in a conventional CDMA system with the same degradation. This ideally represents an almost ninefold increase in the user capacity of the system.

V. Multipath Considerations

In both the analysis and simulation results presented here, we assumed an AWGN channel. If multipath is present and assumed to be known, and can be modeled as L distinct multipath rays, then the following modifications of the above scheme would take place. First, in the description of the received signal, one would replace the modulation pulse shape $p(t)$ with its channel output version, namely $p'(t) = \sum_{l=1}^{L} h_l p(t - \tau_l)$, τ_l are the multipath delays, and h_l the

[6]The optimum values of p_k for the three stages were basically found by a trial-and-error computer search with the only restriction being $p_1 < p_2 < p_3$. This restriction is intuitively clear as previously discussed near the end of Section III.

multipath channel coefficients. Next, the correlator in the receiver would be replaced by an optimal coherent RAKE combiner, and finally, the respreader (multiplication of the tentative decision by the PN code) would be replaced by a circuit imitating the effects of the multipath on the PN code. This circuit should compute

$$\hat{a}_k PN_k'(t) = \sum_{l=1}^{L} h_l PN_k(t-\tau_l)\hat{a}_k. \tag{20}$$

VI. Conclusions

The inclusion of multistage parallel interference cancellation in a CDMA receiver can significantly improve its performance relative to that of a conventional CDMA receiver where no interference cancellation is attempted. A partial interference cancellation philosophy, in which the amount of interference canceled is related to the fidelity of the tentative decisions involved in forming the interference estimate, is in general superior to a brute force philosophy of entirely canceling the interference at each stage. Using a hyperbolic tangent device for making the tentative decisions at the various stages of the cancellation process is superior to using either a hard limiter or linear device. The linear device, on the other hand, has the advantage that the receiver implementation does not require knowledge of the user powers nor does it need carrier synchronization at the various stages. The latter implies that the final data decisions can be performed with a differential (rather than a coherent) detector. The technique is equally applicable to uncoded as well as coded modulations, the latter being discussed in [18]. Finally, the authors wish to alert the readers to an excellent survey article [20] on multiuser detection of CDMA which appeared after our paper was submitted for publication but includes reference to our work as originally presented at the 1995 IEEE Communication Theory Workshop and later reported in [18]. Another survey article that deserves mention, in particular because of its focus on the performance in the presence of multipath propagation, is the work of Duel-Hallen, Holtzman, and Zvonar [21].

Appendix A
Derivation of the Nonlinear MMSE Estimate $\hat{a}_1(k)$

Consider using for $\hat{a}_1(k)$ the nonlinear estimate $E\{a_1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)\}$ which is given by

$$\begin{aligned} &E\{a_1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)\} \\ &\quad = (1)\times p[1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)] + (-1) \\ &\qquad \cdot p[-1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)] \end{aligned} \tag{A.1}$$

with $p[a_1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)]$ the *a posteriori* probability of the user 1's bit (in the transmission interval $0 \le t \le T_b$) given the observations and interference. Using Bayes' rule, this probability can be determined in terms of the conditional probability $p[Y_1, \tilde{a}_1(k-1)|a_1, \hat{I}_1(k)]$ of (11) of the main text as

$$p[a_1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)] = \frac{\Lambda(a_1)q(a_1)}{\Lambda(1)q(1)+\Lambda(-1)q(-1)} \tag{A.2}$$

where

$$\begin{aligned} \Lambda(a_1) &\triangleq p[Y_1, \tilde{a}_1(k-1)|a_1, \hat{I}_1(k)] \\ q(a_1) &= p[a_1, \hat{I}_1(k)]. \end{aligned} \tag{A.3}$$

Evaluating (A.2) at $a_1 = 1$ and $a_1 = -1$ and substituting the results into (A.1) gives

$$E\{a_1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)\} = \frac{\Lambda(1)q(1)-\Lambda(-1)q(-1)}{\Lambda(1)q(1)+\Lambda(-1)q(-1)}. \tag{A.4}$$

Because of the symmetry of the problem, i.e., the equiprobable properties of the data streams, we have that $q(1) = q(-1)$. Hence, (A.4) becomes

$$\hat{a}_1(k) \triangleq E\{a_1|Y_1, \tilde{a}_1(k-1), \hat{I}_1(k)\} = \frac{\Lambda(1)-\Lambda(-1)}{\Lambda(1)+\Lambda(-1)}. \tag{A.5}$$

Referring to (14) of the main text for the evaluation of $\Lambda(a_1)$, we obtain the desired result

$$\begin{aligned} \hat{a}_1(k) &= \frac{C\exp(\alpha_k\lambda_k) - C\exp(-\alpha_k\lambda_k)}{C\exp(\alpha_k\lambda_k) + C\exp(-\alpha_k\lambda_k)} = \tanh(\alpha_k\lambda_k) \\ &= \tanh\Bigg(\sqrt{\frac{2E_{b1}}{N_0}}\left[\frac{\sigma_{1k}^2+\sigma_{2,k-1}^2-2\rho_k\sigma_{1k}\sigma_{2,k-1}}{\sigma_{1k}^2\sigma_{2,k-1}^2(1-\rho_k^2)}\right] \\ &\qquad \cdot \{p_k[Y_1-\hat{I}_1(k)] + (1-p_k)\tilde{a}_1(k-1)\}\Bigg). \end{aligned} \tag{A.6}$$

References

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

[2] ——, "Optimum multiuser asymptotic efficiency," *IEEE Tran. Commun.*, vol. COM-34, pp. 890–897, Sept. 1986.

[3] K. S. Schneider, "Optimum detection of code division multiplexed signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, pp. 181–183, Jan. 1979.

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

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

[6] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 641–649, May 1990.

[7] P. W. Dent, "CDMA subtractive demodulation," US Patent 5 218 619, June 8, 1993.

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

[9] ——, "Near optimum detection in synchronous code-division multiple-access systems," *IEEE Trans. Commun.*, vol. 39, pp. 725–736, May 1991.

[10] 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.

[11] Y. C. Yoon, R. Kohno, and H. Imai, "Cascaded co-channel interference cancelling and diversity combining for spread-spectrum multi-access over multipath fading channels," in *Symp. Information Theory and its Applications (SITA'92)*, Minakami, Japan, Sept. 8–11, 1992.

[12] ——, "A spread-spectrum multi-access system with a cascade of co-channel interference cancellers for multipath fading channels," in *Int. Symp. Spectrum Techniques and Applications (ISSSTA'92)*, Yokohama, Japan, Nov. 29–Dec. 2, 1992.

[13] ———, "A spread-spectrum multiaccess system with cochannel interference cancellation," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 1067–1075, Sept. 1993.

[14] M. Kawabe, T. Kato, A. Kawahashi, T. Sato, and A. Fukasawa, "Advanced CDMA scheme based on interference cancellation," in *Proc. 43rd Annu. IEEE Vehicular Technology Conf.*, May 18–20, 1993, pp. 448–451.

[15] S. Verdu and H. V. Poor, "Abstract dynamic programming models under commutativity conditions," *SIAM J. Control and Optimization*, vol. 25, pp. 990–1006, July 1987.

[16] H. V. Poor, "On parameter estimation in DS/SSMA formats," in *Advances in Communications and Signal Processing*, A. Porter and S. C. Kak, Eds. New York: Springer-Verlag, 1989, pp. 59–70.

[17] C. L. Weber, G. K. Huth, and B. H. Batson, "Performance considerations of code division multiple-access systems," *IEEE Trans. Veh. Technol.*, vol. VT-30, pp. 3–10, Feb. 1981.

[18] D. Divsalar and M. K. Simon, "Improved CDMA performance using parallel interference cancellation," JPL Publication 95-21, Oct. 1995.

[19] 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.

[20] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, pp. 124–136, Oct. 1996.

[21] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Personal Commun.*, pp. 46–58, Apr. 1995.

Dariush Divsalar (S'76–M'78–SM'90–F'97), for photograph and biography, see this issue, p. 210.

Marvin K. Simon (S'60–M'66–SM'75–F'78), for photograph and biography, see this issue, p. 210.

Dan Raphaeli (M'92), for photograph and biography, see p. 40 of the January 1998 issue of this TRANSACTIONS.