Linear versus Non-linear Interference Cancellation

R. Michael Buehrer, Steven P. Nicoloso, and Sridhar Gollamudi

Abstract: **In this paper we compare linear and non-linear interference cancellation for systems employing code division multiple access (CDMA) techniques. Specifically, we examine linear and non-linear parallel interference cancellation (also called multistage cancellation) in relationship to other multiuser detection algorithms. We show the explicit relationship between parallel interference cancellation and the decorrelator (or direct matrix inversion). This comparison gives insight into the performance of parallel interference cancellation (PIC) and leads to better approaches. We also show that non-linear PIC approaches with explicit channel estimation can provide performance improvement over linear PIC, especially when using soft non-linear symbol estimates. The application of interference cancellation to non-linear modulation techniques is also presented along with a discussion on minimum mean-squared error (MMSE) symbol estimation techniques. These are shown to further improve the performance of parallel cancellation.**

Index Terms: **Parallel Interference Cancellation, Multiuser Detection, Partial Interference Cancellation, Minimum Mean-squared Error (MMSE) Symbol Estimation, Linear CDMA Interference Suppression, Multistage Receiver.**

I. INTRODUCTION

Multiuser detection has been proposed as a means of improving the capacity of code division multiple access (CDMA) systems relative to the conventional receiver [1]–[4]. The conventional receiver for CDMA is a multi-finger Rake receiver using filters matched to a particular spreading code on each finger, equal gain (or maximal ratio) combining, and square-law (or coherent) symbol detection. Unlike the conventional receiver, multiuser detectors attempt to detect all CDMA signals simultaneously, taking into account the structure of the received signal. The conventional receiver treats multi-access interference (MAI) as additive white Gaussian noise (AWGN). This is a reasonable approach due to the fact that CDMA interference is comprised of contributions from many independent interferers, and during despreading, the contribution of each signal is randomized by the correlation of the various spreading codes. As a result, MAI will appear as a single additive noise source. It is thus the *average* interference that limits CDMA system performance, rather than the *worst case* interference as with orthogonal access systems. However, the MAI is not true AWGN since it is predictable or at least has known properties that can be exploited. The exploitation of this information is the idea behind multiuser reception, allowing greatly enhanced link performance.

The terms multiuser detection and interference cancellation are often used interchangeably. For the sake of this discussion, we reserve the term interference cancellation for multiuser detection techniques which attempt to estimate the interference, regenerate it, and cancel it through subtractive means. Multiuser detection however, will refer to the more general detection algorithms which use some knowledge of the interference to improve performance. Linear interference cancellation algorithms are techniques which attempt to regenerate and cancel interference in such a way that the final decision statistic is a linear function of the matched filter outputs. Linear multiuser detection algorithms also create decision statistics which are a linear combination of the matched filter outputs, but are not explicitly attempting direct subtractive cancellation. We will show that these techniques are intimately related. Non-linear techniques are any techniques which use non-linear functions of the matched filter outputs.

Parallel interference cancellation (PIC), also known as multistage detection, is one form of interference cancellation which is applicable to a wide variety of CDMA system designs (e.g., long or short spreading codes, linear or non-linear modulation). Significant research has been done in this area over the last several years on the both linear and non-linear PIC [5]–[8]. In this work, we first examine linear PIC and show how it relates to other linear multiuser detection algorithms including the decorrelating [9] and MMSE [10] receivers as well as a linear form of successive interference cancellation [11]. We then discuss non-linear PIC and compare its performance with linear PIC as well as other non-linear interference cancellation algorithms. Non-linear PIC has the requirement that it requires channel estimates for cancellation purposes while linear PIC does not. However, this requirement does provide potential performance improvements, particularly when non-linear symbol estimation is used.

We first present the CDMA system model in section II upon which all discussion is based. In section III, we put PIC into a context of linear multiuser detection algorithms by first considering linear interference cancellation. Specifically, we show the mathematical relationship between PIC with other linear multiuser detection, most importantly the decorrelator and linear MMSE receiver. This relationship provides insight into the ultimate performance and limitations of linear PIC. Decision statistic bias motivates a discussion of partial interference cancellation which is shown to mitigate the bias problem. Limitations of PIC, however, can also be mitigated in part by simply removing the linear constraint on PIC and considering non-linear methods of PIC. This we do in section IV.

Section IV considers non-linear PIC and compares its per-

Manuscript received August 31, 1998; approved for publication by Elvino Sousa, Division I Editor, May 13, 1999.

R. M. Buehrer and S. P. Nicoloso are with Lucent Technologies, Wireless Technology Laboratory, Whippany, NJ, e-mail: mbuehrer@Bell-labs.com and spn@lucent.com.

S. Gollamudi is currently with Department of Electrical Engineering, University of Notre Dame, South Bend, IN.

1229-2370/99/$10.00

formance and complexity to linear cancellation as well as other non-linear multiuser detection techniques. Relaxation of the linear constraint offers the important ability for a PIC receiver to decouple channel and symbol estimation, allowing them to be considered separately, and motivates a description of MMSE signal reconstruction for PIC. Given the MMSE criterion, we examine the non-linear MMSE (optimal) symbol estimators for linear modulation (e.g., BPSK), and for non-linear modulation derived in [12]. The MMSE symbol estimate for linear modulation is shown to provide some performance gain over a hard-limited non-linear estimate. With non-linear modulation, viz., M-ary orthogonal modulation, it is shown to provide much more substantial gains. Issues arising from decoupling the channel estimation block are also considered. Non-linear PIC is shown not only to out-perform linear multiuser detection techniques (including linear PIC) in most circumstances, but also to be more widely applicable to actual CDMA systems.

II. SYSTEM MODEL

To clarify the representation developed used in this paper we begin with a mathematical model of the CDMA system under consideration. The received CDMA signal is represented by

$$r(t) = \sum_{k=1}^{K} \sqrt{P_k} b_k(t-\tau_k) s_k(t-\tau_k) e^{j\theta_k} + n_o(t), \quad (1)$$

where K users are independently transmitting bi-phase modulated signals, P_k is the power received[1], $b_k(t)$ is the data signal with symbol period T_b, $s_k(t)$ is the spreading signal with chip width T_c, and τ_k is the relative delay of the kth user. The complex Gaussian noise process $n_o(t)$ has power σ^2, and the spreading gain is defined to be $N = \frac{T_b}{T_c}$. Note that we have assumed perfect timing acquisition (i.e., $\hat{\tau}_k = \tau_k$) and phase distortion due to the channel and imperfect down-conversion is constant over the observation interval.

The phase-compensated output of a filter matched to the kth user's spreading waveform, after the ith bit interval (assuming perfect carrier and PN code phase tracking) is represented by

$$y_{i,k} = \Re\left\{ \frac{1}{T_b} \int_{(i-1)T_b+\tau_k}^{iT_b+\tau_k} r(t) s_k^*(t-\tau_k-iT_b) e^{-j\hat{\theta}_k} dt \right\}, \quad (2)$$

where $\Re(\cdot)$ is the real operator. This leads to a vector representation of matched filter outputs corresponding to the ith bit interval,

$$\mathbf{y}_i = \mathbf{R}(-1)\mathbf{A}_{i-1}\mathbf{b}_{i-1} + \mathbf{R}(0)\mathbf{A}_i\mathbf{b}_i + \mathbf{R}(1)\mathbf{A}_{i+1}\mathbf{b}_{i+1} + \mathbf{n}_i, \quad (3)$$

where $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,K}]^T$, $\mathbf{R}(m)$, $m \in \{-1, 0, 1\}$, is a $K \times K$ matrix that represents the partial correlation between users over the mth relative bit interval (i.e., between the ith bit of user j and the $(i+m)$th bit of user k)[13], $\mathbf{A}_i$ is a diagonal matrix with vector $[a_{i,1}, a_{i,2}, \ldots, a_{i,K}]^T$ along the diagonal where $a_{i,k} = \sqrt{P_k(t)}, t = iT_b$ is the received amplitude of the kth user, $(\cdot)^T$ is the transpose operator, $\mathbf{b}_i = [b_{i,1}, b_{i,2}, \ldots, b_{i,K}]^T$, and $\mathbf{n}_i = [n_{i,1}, n_{i,2}, \ldots, n_{i,K}]^T$ is a vector of filtered noise samples. Note also that equation (3) assumes perfect phase information, i.e., $\hat{\theta} = \theta$. This assumption is made for notational convenience in the next section. However, we will drop this assumption when we discuss linear interference cancellation.

Given a sequence of N_b data symbols from each of K users, we represent the sequence of matched filter outputs by

$$\mathbf{y} = \mathcal{R}\mathcal{A}\mathbf{b} + \mathbf{n}, \quad (4)$$

where $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots \mathbf{y}_{N_b}^T]^T$, $\mathbf{b}$ and $\mathbf{n}$ are similarly defined, $\mathcal{A}$ is a $KN_b \times KN_b$ diagonal matrix with $[\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_{N_b}]^T$ along the diagonal, and

$$\mathcal{R} = \begin{pmatrix} \mathbf{R}(0) & \mathbf{R}(-1) & 0 & \cdots & 0 \\ \mathbf{R}(1) & \mathbf{R}(0) & \mathbf{R}(-1) & & \vdots \\ 0 & \mathbf{R}(1) & \mathbf{R}(0) & \ddots & 0 \\ \vdots & & \ddots & \ddots & \mathbf{R}(-1) \\ 0 & \cdots & 0 & \mathbf{R}(1) & \mathbf{R}(0) \end{pmatrix}. \quad (5)$$

Note that while the definitions of $\mathbf{R}$ and $\mathcal{R}$ imply short, i.e., non-time-varying or "code-on-pulse," spreading sequences, techniques in the proceeding discussion do not require such an assumption. This is intended merely to simplify notation.

Finally, there will be times when we will not wish to assume that phase compensation is performed on the matched filter outputs. Thus, we define

$$\tilde{y}_{i,k} = \frac{1}{T_b} \int_{(i-1)T_b+\tau_k}^{iT_b+\tau_k} r(t) s_k^*(t-\tau_k-iT_b) dt \quad (6)$$

or $\tilde{y}_{i,k} = y_{i,k} e^{j\theta_k}$. Additionally, we define $\tilde{a}_{i,k} = a_{i,k} e^{j\theta_k}$ and $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{A}}$ represent the vector versions of $\tilde{y}_{i,k}$ and $\tilde{a}_{i,k}$ in a manner similar to $\mathbf{y}$ and $\mathcal{A}$.

III. LINEAR CANCELLATION

Multiuser detection schemes can be grouped into two broad categories, linear and non-linear [2]. The former category contains the decorrelating receiver [9] and the linear MMSE receiver [9], [10]. These receivers use knowledge of the spreading codes of all users (with the exception of adaptive MMSE receivers [14]) to create a linear transformation that projects each signal onto a subspace dependent upon design criteria. The decorrelator removes the effects of MAI by projecting the signal of interest onto a subspace which is orthogonal to the entire interference subspace. Since this subspace is not in the same direction as the desired signal, the decorrelator suffers a loss of desired signal energy and thus the performance versus thermal noise degrades. However, the receiver does eliminate the effect of interference. The MMSE receiver alleviates this drawback by projecting the desired signal in a direction which minimizes the combined effect of MAI and thermal noise. An additional benefit of the MMSE structure is that it can be implemented adaptively and blindly [14].

[1] Note that for case of analysis we have assumed throughout that P_k is constant over the observation interval, i.e., we have ignored the effect of multipath fading.

In general, a linear detector (assuming BPSK signaling) is one that makes decisions based on

$$\begin{aligned}\hat{\mathbf{b}} &= \operatorname{sgn}(\mathbf{Ty}) \\ &= \operatorname{sgn}(\mathbf{T}(\mathcal{R}\mathcal{A}\mathbf{b}+\mathbf{n})),\end{aligned} \tag{7}$$

where $\mathbf{T}$ is a linear operator on $\mathbf{y}$. Using near-far resistance[2] to drive suboptimal[3] receiver development, Lupas showed that a linear receiver can obtain the near-far resistance of the optimal receiver [13]. The linear detector that accomplishes this is the decorrelating detector. Viewing the matched filter receiver outputs as the output of a linear system, the decorrelating detector simply performs a direct matrix inverse of system transfer function $\mathcal{R}$. In other words, $\mathbf{T}=\mathcal{R}^{-1}$ which leads to

$$\begin{aligned}\hat{\mathbf{b}} &= \operatorname{sgn}(\mathcal{R}^{-1}\mathbf{y}) \\ &= \operatorname{sgn}(\mathcal{A}\mathbf{b}+\mathcal{R}^{-1}\mathbf{n}).\end{aligned} \tag{8}$$

Since $\mathcal{R}$ as defined is simply the normalized cross-correlation between the users' spreading codes, the decorrelator is independent of the received user energies. Because of this, the decorrelator is the optimal linear receiver for unknown signal energies[4] [13]. This obviates the need for estimates of the channel to remove the effects of interference, although phase estimation is obviously still required for coherent detection. Additionally, (8) clearly shows that the data estimate is independent of the interfering powers thus providing near-far resistance.

There are, however, two main disadvantages of the decorrelating receiver. The first is the need to calculate the inverse of the cross-correlation matrix in order to obtain the decorrelation coefficients. The second is that in high noise situations, receiver performance degrades due to the correlation introduced in the noise. This noise enhancement is analogous to that seen with a zero-forcing equalizer for inter-symbol interference (ISI), where the application of a channel inverse creates correlated noise samples. With the decorrelator, this effect increases noise power in proportion to the cross-correlation between users. To show this, we define $\Sigma_y = \operatorname{var}[\mathbf{y}\,|\mathbf{b},\mathcal{A}]$ and

$$\begin{aligned}\Sigma_y &= \mathrm{E}\left[\mathbf{y}\mathbf{y}^{\mathrm{T}}\right]-\mathrm{E}\left[\mathbf{y}\right]\mathrm{E}\left[\mathbf{y}^{\mathrm{T}}\right] \\ &= \mathrm{E}\left[(\mathcal{A}\mathbf{b}+\mathcal{R}^{-1}\mathbf{n})(\mathcal{A}\mathbf{b}+\mathcal{R}^{-1}\mathbf{n})^{\mathrm{T}}\right]-(\mathcal{A}\mathbf{b})(\mathcal{A}\mathbf{b})^{\mathrm{T}} \\ &= \mathcal{A}\mathbf{b}\mathrm{E}[\mathbf{n}^{\mathrm{T}}](\mathcal{R}^{-1})^{\mathrm{T}}+\mathcal{R}^{-1}\mathrm{E}[\mathbf{n}]\mathbf{b}^{\mathrm{T}}\mathcal{A}^{\mathrm{T}} \\ &\quad +\mathcal{R}^{-1}\mathrm{E}\left[\mathbf{n}\mathbf{n}^{\mathrm{T}}\right](\mathcal{R}^{-1})^{\mathrm{T}} \\ &= \sigma^2\mathcal{R}^{-1},\end{aligned} \tag{9}$$

[2]Please see [15] for a definition and discussion of near-far resistance.

[3]The optimal receiver is the general maximum likelihood receiver, i.e., the receiver which chooses $\hat{\mathbf{b}}$ as

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}\in\{-1,1\}^K}\int_0^T [r(t)-S(\mathbf{b})]^2\,dt,$$

where $S(\mathbf{b})=\sum_{k=1}^{K}\sum_{i=-M}^{M}\sqrt{P_k}b_k[i]p_T(t-iT-\tau_k)s_k(t-\tau_k)$.

[4]That is it chooses

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}\in\{-1,1\}^K}\min_{P_k,\,k=1,\ldots,K}\int_0^T [r(t)-S(\mathbf{b})]^2\,dt.$$

where we have used $\mathrm{E}[\mathbf{n}]=0$ and $\mathrm{E}[\mathbf{n}\mathbf{n}^{\mathrm{T}}]=\sigma^2\mathcal{R}$, and σ^2 is the power of the AWGN at the receiver. Thus the decorrelation process, although removing MAI, enhances the noise.

In [16], linear detectors based on MMSE are proposed. Such a detector attempts to minimize

$$E\left[(\mathbf{b}-\hat{\mathbf{b}})^{\mathrm{T}}(\mathbf{b}-\hat{\mathbf{b}})\right], \tag{10}$$

where $E[\cdot]$ is the expectation operator and $\hat{\mathbf{b}}=\operatorname{sgn}(\mathbf{Ty})$ for some linear transformation $\mathbf{T}$. The linear transformation $\mathbf{T}$ which attains the minimum value is

$$\mathbf{T}=\left(\mathcal{R}+\frac{N_o}{2\mathcal{A}^2}\right)^{-1}. \tag{11}$$

This leads to the decision rule

$$\hat{\mathbf{b}}=\operatorname{sgn}\left[\left(\mathcal{R}+\frac{N_o}{2\mathcal{A}^2}\right)^{-1}\mathbf{y}\right]. \tag{12}$$

From (11) we can see that for $N_o\simeq 0$ (i.e., low noise situations) the MMSE and decorrelator are identical. In the opposite extreme (N_o extremely large), the MMSE detector reduces to the conventional receiver. Like the decorrelating receiver, the MMSE receiver obtains optimal near-far resistance, however, unlike the decorrelator it requires knowledge of the received user energies.

A. Relationship Between PIC and the Decorrelator

Generally speaking, parallel interference cancellation uses estimates of the channel along with tentative symbol estimates to iteratively remove the effect of interference on each desired signal. The decision rule at each stage (iteration) for PIC with BPSK signaling is represented in matrix notation by

$$\begin{aligned}\hat{\mathbf{b}} &= \operatorname{sgn}\left[\mathbf{z}\hat{\mathcal{A}}^*\right] \\ &= \operatorname{sgn}\left[\left(\mathcal{R}\tilde{\mathcal{A}}\tilde{\mathbf{b}}+\mathbf{n}-\mathcal{P}\hat{\mathcal{A}}\tilde{\mathbf{b}}\right)\hat{\mathcal{A}}^*\right],\end{aligned} \tag{13}$$

where $\mathbf{z}$ is the decision statistic, $\mathcal{P}=\mathcal{R}-\mathcal{I}$, $\mathcal{I}$ is the identity matrix, $\hat{\mathcal{A}}$ is an estimate of the diagonal matrix of complex amplitudes $\tilde{\mathcal{A}}$, $\tilde{\mathbf{b}}$ is the estimated data symbols. For linear soft decision cancellation we set the initial combined estimate $\hat{\mathcal{A}}\tilde{\mathbf{b}}$ equal to the complex correlator outputs $\tilde{\mathbf{y}}$ resulting in the decision statistic after one stage of cancellation[5]

$$\begin{aligned}\mathbf{z}^{(2)} &= \mathcal{R}\tilde{\mathcal{A}}\tilde{\mathbf{b}}+\mathbf{n}-\mathcal{P}\tilde{\mathbf{y}} \\ &= [\mathcal{I}-\mathcal{P}]\,\mathcal{R}\tilde{\mathcal{A}}\tilde{\mathbf{b}}+[\mathcal{I}-\mathcal{P}]\,\mathbf{n}.\end{aligned} \tag{14}$$

Subsequently using $\hat{\mathcal{A}}\tilde{\mathbf{b}}=\mathbf{z}^{(2)}$ in the second stage stage of cancellation results in

$$\mathbf{z}^{(3)}=\left[\mathcal{I}-\mathcal{P}+\mathcal{P}^2\right]\mathcal{R}\tilde{\mathcal{A}}\tilde{\mathbf{b}}+\left[\mathcal{I}-\mathcal{P}+\mathcal{P}^2\right]\mathbf{n}. \tag{15}$$

Further, it can be shown that in general

$$\mathbf{z}^{(S)}=\mathcal{T}^{(S)}\left(\mathcal{R}\tilde{\mathcal{A}}\tilde{\mathbf{b}}+\mathbf{n}\right), \tag{16}$$

[5]Note that we now drop the assumption of channel phase knowledge being incorporated in the matched filter (correlator) outputs and use $\tilde{\mathbf{y}}$ rather than $\mathbf{y}$.

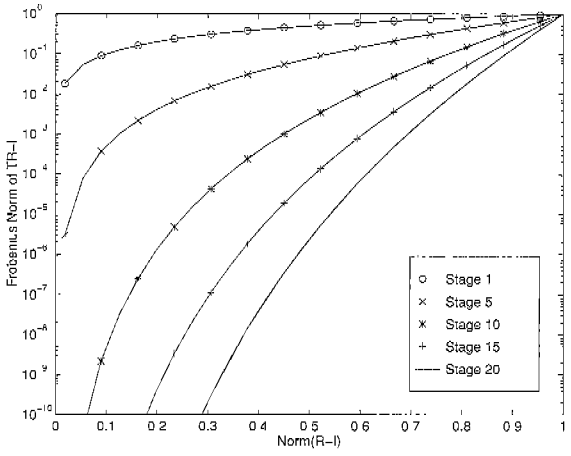

Fig. 1. The degree to which S-stage linear parallel interference cancellation approximates the decorrelating receiver as the average normalized cross-correlation increases. (synchronous system, stage 1: conventional receiver, norm(TR-I) represents the residual cross-correlation after cancellation.)

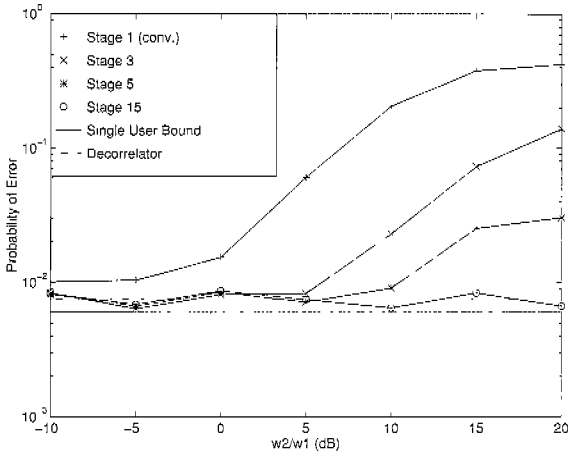

Fig. 2. The probability of error for S-stage parallel cancellation vs. interference power for several values of S as the ratio between the interfering power and the desired power increases. (stage 1: conventional receiver, $E_b/N_o = 5$dB, $N = 15$, $P_3 = P_1$.)

where $\mathcal{T}^{(S)}$ is a linear transformation at stage S determined by

$$\mathcal{T}^{(S)} = \sum_{s=0}^{S-1} (-1)^s \mathcal{P}^s. \tag{17}$$

Now, if we allow the number of stages to approach infinity we obtain

$$\begin{aligned} \lim_{S\to\infty} \mathcal{T}^{(S)} &= \lim_{S\to\infty} \sum_{s=0}^{S-1} (-1)^s \mathcal{P}^s \\ &= (\mathcal{P} + \mathcal{I})^{-1} \\ &= \mathcal{R}^{-1} \end{aligned} \tag{18}$$

provided $||\mathcal{P}||_2 < 1$.[6] Thus, in the limit the PIC receiver has a decision statistic equal to

$$\lim_{S\to\infty} \mathbf{z}^{(S)} = \tilde{\mathcal{A}}\mathbf{b} + \mathcal{R}^{-1}\mathbf{n} \tag{19}$$

which is equivalent to the decorrelating receiver, the same conclusion reached in [17] and [18]. The number of stages necessary for a good approximation depends on the total cross-correlation between users (i.e., system loading) as will be shown.

To investigate the number of stages needed for convergence, simulations were run for a synchronous system with various loading levels, random spreading codes and a spreading gain of N=64. The results were averaged over 1000 simulation runs. Fig. 1 shows the Frobenius norm of the approximation error, $||\mathcal{T}^{(S)}\mathcal{R} - I||_F$ versus the total normalized cross-correlation measured by $||\mathcal{R} - \mathcal{I}||_2$ for a synchronous system. The error can be interpreted as the residual cross-correlation after the transformation (i.e., after cancellation), since it will depend on the off-diagonal elements. For perfect decorrelation, $\mathcal{T} = \mathcal{R}^{-1}$ and $||\mathcal{T}\mathcal{R} - I||_F = 0$. This plot shows that as the total system cross-correlation increases (meaning that either the system loading increases or the spreading gain decreases), more stages are required to obtain the same residual interference level. This error is most meaningful in near-far situations where small amounts of cross-correlation can result in significant interference levels. In other words, as the system loading increases, more stages are required for accurate approximation of the decorrelating receiver. As the total normalized cross-correlation approaches unity, the linear PIC receiver will not converge to the decorrelator. This will be discussed further in section III-B.1.

While the previous results provided insight into convergence, they do not translate directly into BER performance. Thus, simulations were run to determine the probability of error performance of S-stage parallel interference cancellation in the presence of strong interference as the number of stages or iterations increases. Fig. 2 presents simulation results for the case of 3 users, a spreading gain $N = 15$ (random codes), and asynchronous reception. Signals 1 and 3 are perfectly power controlled to provide a received $E_b/N_o = 5$dB while the received strength of signal 2 is allowed to vary from -10dB to 20dB relative to signals 1 and 3. The average probability of error for signal 1 is plotted for several stages along with the conventional receiver (stage 1), the decorrelator, and the single user bound. As we would expect, the conventional receiver fails in the presence of significantly strong interference. As S increases, the robustness to interference strength is increased. For $S = 15$ we see that performance is independent of interference power over the range of interest and equivalent to the decorrelator. Note that the performance is slightly degraded compared to the single user bound due to the noise enhancement suggested by equation (9) [13].

[6] Note that this is true for any matrix norm. The requirements for convergence are examined in more detail in section III-B.1.

B. Matrix Interpretation of PIC

The previous section showed that linear PIC will approach the decorrelating receiver for a sufficient number of stages and given certain conditions on the cross-correlation between spreading codes. This leads to the investigation of PIC in terms of solving linear systems, i.e., using a matrix interpretation[7]. In the following section we show that linear PIC is equivalent to the Jacobi iteration method of solving linear systems. This leads one to investigate other methods of solving linear systems as forms of interference cancellation. Specifically, we formulate Gauss-Seidel iterations and gradient search methods in terms of interference cancellation in this section.

B.1 Jacobi Iterations

One method of solving a linear set of equations $\mathbf{Mx} = \mathbf{c}$ is the method of Jacobi iterations. If we rearrange the linear set of equations it can be shown that

$$x_i = \frac{\left(c_i - \sum_{j \neq i} m_{ij} x_j\right)}{m_{ii}}, \tag{20}$$

where m_{ij} is that element in the ith row and jth column of $\mathbf{M}$. Given an initial estimate of x_i, denoted $\hat{x}_i^{(0)}$, we can iteratively calculate $\hat{x}_i$ by

$$\hat{x}_i^{(k)} = \frac{\left(c_i - \sum_{j \neq i} m_{ij} \hat{x}_j^{(k-1)}\right)}{m_{ii}}. \tag{21}$$

At convergence

$$\hat{\mathbf{x}} = \mathbf{D}^{-1}\left(\mathbf{c} - (\mathbf{M} - \mathbf{D})\hat{\mathbf{x}}\right), \tag{22}$$

where $\mathbf{D}$ is a diagonal matrix with the diagonal elements of $\mathbf{M}$. Solving for $\hat{\mathbf{x}}$ gives

$$\hat{\mathbf{x}} = \mathbf{M}^{-1}\mathbf{c}. \tag{23}$$

Thus, provided that the iteration converges, we will arrive at the linear system solution. In terms of BPSK modulated CDMA[8], we consider the system

$$\tilde{\mathbf{y}} = \mathbf{R}\tilde{\mathbf{A}}\mathbf{b} + \mathbf{n}, \tag{24}$$

where $\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_i, \forall i$, $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_i, \forall i$, $\mathbf{R} = \mathbf{R}(\mathbf{0})$, and $\mathbf{n} = \mathbf{n_i}$ are the synchronous equivalents, i.e., $\tau_k = 0, \forall k$, of the variables defined in section II. Consider the problem of solving for $\mathbf{R}^{-1}\tilde{\mathbf{y}}$ by using the iteration in (21). This leads to the decision statistic

$$z_i^{(1)} = \tilde{y}_i - \sum_{j \neq i} \rho_{ij} \tilde{y}_j, \tag{25}$$

where ρ_{ij} are the elements of $\mathbf{R}$, $\rho_{ii} = 1\ \forall i$, and $z_i^{(0)} = \tilde{y}_i$. In matrix form we have

$$\mathbf{z^{(1)}} = \tilde{\mathbf{y}} - (\mathbf{R} - \mathbf{I})\tilde{\mathbf{y}}. \tag{26}$$

[7]At the time of submission it was brought to our attention that [19] also shows that parallel cancellation can be formulated in terms of matrix algebra.

[8]Here we assume synchronism for notational convenience, but it is not required.

It can be shown that

$$\mathbf{z}^{(k)} = \sum_{i=0}^{k} (-1)^i (\mathbf{R} - \mathbf{I})^i \tilde{\mathbf{y}}, \tag{27}$$

which is identical to the linear PIC receiver described in section III-A. Thus, when using matched filter outputs directly as interference estimates, PIC is linear and identical to the Jacobi iteration method of solving a linear system.

This equivalence provides some insight into the performance of linear PIC using soft estimates. First, we can gain an understanding of the applicability of PIC implemented in this manner by examining the convergence properties of the Jacobi iteration method. Rewriting equation (27) in recursive form we can state the PIC method as

$$\mathbf{z}^{(k+1)} = -\mathbf{P}\mathbf{z}^{(k)} + \tilde{\mathbf{y}}, \tag{28}$$

where $\mathbf{P} = \mathbf{R} - \mathbf{I}$. Now at convergence $\mathbf{z} = \mathbf{R}^{-1}\mathbf{y}$. However, the iterations will converge iff $\gamma(\mathbf{P}) < 1$ where $\gamma(\mathbf{P})$ is the spectral radius of $\mathbf{P}$, i.e.,

$$\gamma(\mathbf{P}) = \max\{|\lambda| : \lambda \in \Lambda(\mathbf{P})\} \tag{29}$$

and $\Lambda(\mathbf{P})$ are the eigenvalues of $\mathbf{P}$ [20]. Thus, the solution will only converge to the direct matrix inversion (i.e., the decorrelator) if the matrix $\mathbf{R}$ is diagonally dominant, where diagonal dominance is defined as

$$\sum_{\substack{j=1 \\ j \neq i}}^{K} |\ \rho_{ij}\ | < |\ \rho_{ii}\ | = 1 \qquad i = 1, 2, ..., K. \tag{30}$$

In other words, PIC formulated in this way will converge as the number of stages increases if the total cross-correlation between any spreading code and all other codes is less than unity. This is a fairly strong requirement, especially with random spreading codes, and indicates that this implementation of PIC will not likely be useful for even a moderate number of users.

Another interesting note about linear PIC from this analysis is that channel knowledge is not required for cancellation[9]. Phase information is required for *detection* in equation (13) (and Rake combining in a multipath scenario), but for cancellation purposes no channel knowledge is necessary. This can be seen in equation (27) and is because only the complex matched filter outputs are used along with knowledge of $\mathbf{R}$. Additionally, if DPSK modulation were used no channel knowledge would be required at all since $\hat{\mathbf{b}} = \Re\{\mathbf{z}[i]\mathbf{z}[i-1]^*\}$.

B.2 Gauss-Seidel Iterations

The previous investigation leads one to consider other matrix inversion (or linear system) techniques. One technique that is similar to the Jacobi method, but has faster and more global convergence is Gauss-Seidel iteration. Gauss-Seidel iterations are defined by

$$\hat{x}_i^{(k)} = \frac{\left(c_i - \sum_{j=1}^{i-1} m_{ij} \hat{x}_j^{(k)} - \sum_{j=i+1}^{K} m_{ij} \hat{x}_j^{(k-1)}\right)}{m_{ii}}, \tag{31}$$

[9]By channel information we are referring to knowledge of the random amplitude/phase distortion caused by the channel. Timing information is still necessary.

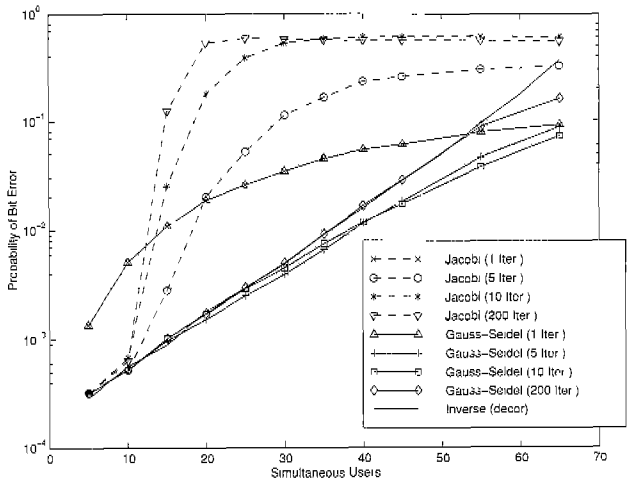

Fig. 3. Performance of Jacobi (JIC) and Gauss-Seidel (GSIC) based cancellation techniques in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, random phase - direct inverse (decorrelator) shown for comparison.)

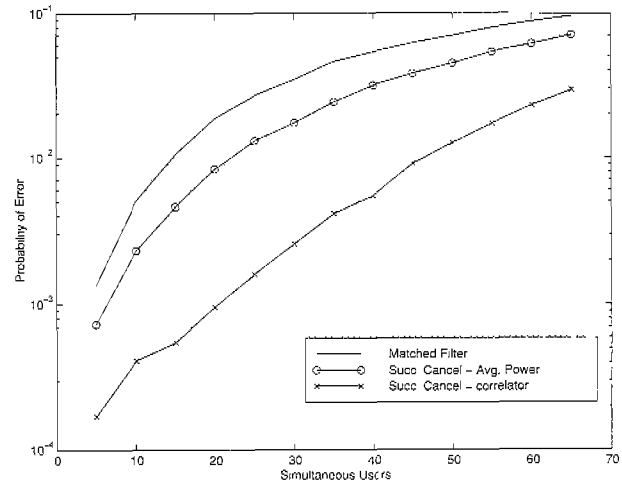

Fig. 4. Performance of linear successive interference cancellation using one symbol estimates for cancellation in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, random phase; for Avg. Powers, ordering done on average power basis; for correlator, ordering done on basis of correlator outputs after each cancellation.)

where we now use the latest information in the calculation of the current estimate. Putting it in terms of BPSK modulated direct sequence CDMA (DS-CDMA), we form a decision statistic of the ith user at the kth stage as

$$z_i^{(k)} = \tilde{y}_i - \sum_{j=1}^{i-1} \rho_{i,j} z_j^{(k)} - \sum_{j=i+1}^{K} \rho_{i,j} z_j^{(k-1)}. \tag{32}$$

We can also state the method in matrix form to show its similarity to Jacobi iterations:

$$\mathbf{z}^{(k+1)} = -\mathbf{G}\mathbf{z}^{(k)} + \tilde{\mathbf{y}}, \tag{33}$$

where $\mathbf{G} = (\mathbf{R} - \mathbf{U})^{-1}\mathbf{U}$ and

$$\mathbf{U} = \begin{pmatrix} 0 & \rho_{12} & \cdots & \cdots & \rho_{1K} \\ 0 & 0 & \cdots & & \vdots \\ 0 & 0 & \ddots & & \rho_{K-2,K} \\ \vdots & & & & \rho_{K-1,K} \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{34}$$

In a method analogous to the above discussion, it can be shown that this method converges to $\mathbf{z} = \mathbf{R}^{-1}\tilde{\mathbf{y}}$ iff $\Lambda(\mathbf{G}) < 1$. It can be shown that this condition is met if $\mathbf{R}$ is symmetric and positive definite. Thus, this method will converge under more general conditions than the Jacobi method.

This technique is similar to linear successive interference cancellation (SIC) (see [11], [21]) with the following exceptions: (1) cancellation using Gauss-Seidel iteration is performed multiple times for each user whereas it usually performed only once with SIC[10]; and more importantly (2) the Gauss-Seidel iteration scheme does not require ordering to improve reliability at each cancellation. Thus, SIC is formulated as

[10]One exception is [22].

$$z_i = \tilde{y}_i - \sum_{j=1}^{i-1} z_j \rho_{ij}, \tag{35}$$

where the order of cancellation is based optimally on reliability but can also be based on average powers [21]. The former ordering method provides superior performance but requires reordering after each cancellation.

Fig. 3 presents the simulated performance of interference cancellation based on Jacobi iterations (JIC) and Gauss-Seidel iterations (GSIC) for a synchronous system in an AWGN channel with $N = 64$, E_b/N_o=8dB, random spreading codes and random phases. Several things can be noted. First we can see that JIC performs very poorly even at moderate loading as predicted. However, GSIC performs very well and converges to the decorrelator in all cases. This is not surprising since the Gauss-Seidel algorithm is known to converge for any initial vector provided the system matrix is symmetric and positive definite which is the case of $\mathbf{R}$. A second item to note is that since the decorrelator does not provide the best performance of any linear detector (the MMSE receiver generally provides better performance), GSIC actually has better performance prior to convergence in some high loading cases. This will be discussed more later.

As a comparison, Fig. 4 presents the performance of SIC implemented with linear cancellation estimates as given in (35). Again the system is synchronous with an AWGN channel, $N = 64$, E_b/N_o=8dB, random spreading codes and random phases. Two ordering techniques are used: (1) ordering based on average powers and (2) ordering based on reliability. As expected, reliability ordering outperforms ordering based on average power. Additionally, we can see that SIC with reliability ordering outperforms GSIC shown in Fig. 3. As points of comparison, SIC with reliability ordering achieves a BER of 10^{-3} with a loading of 20 users and a BER of approximately 10^{-2} at a loading of 50

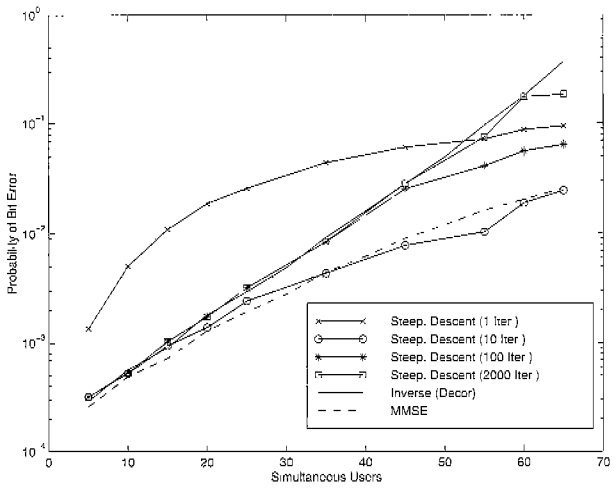

Fig. 5. Performance of steepest descent gradient algorithm in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, random phase; direct inverse (decorrelator) and MMSE algorithms shown for comparison.)

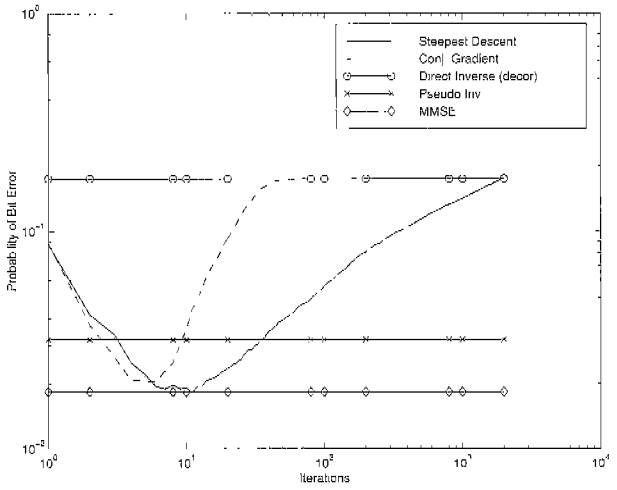

Fig. 6. Performance of steepest descent and conjugate gradient algorithms vs. the number of stages of cancellation in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, $K = 60$ users, random phase; direct inverse (decorrelator) and MMSE algorithms shown for comparison.)

users. GSIC, however, achieves $2 \cdot 10^{-3}$ and $3 \cdot 10^{-2}$ BERs at loadings of 20 and 50 users respectively.

B.3 Gradient Search Methods

A third common method of solving linear systems are the gradient search techniques. It can be shown that the equation

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^H\mathbf{M}\mathbf{x} - \mathbf{x}^H\mathbf{c} \tag{36}$$

is minimized when $\mathbf{x} = \mathbf{M}^{-1}\mathbf{c}$ where $\mathbf{M}$ is real and positive-definite. Thus the minimization of (36) is equivalent to solving $\mathbf{M}\mathbf{x} = \mathbf{c}$. There are two common methods of minimizing (36): the method of steepest descent and the method of conjugate gradients[11]. Both techniques use the negative gradient of the current estimate to choose a new direction for each iteration. However, the method of steepest descent uses the negative gradient directly as the new search direction. The cost function (36) is then minimized in that direction to establish a new estimate. This new estimate is then used to calculate the next gradient. The process continues until a minimum error is calculated or the process reaches a fixed number of iterations. The shortcomings of this technique are well known [20]. Since the method is restricted to search in directions orthogonal to the previous estimate, convergence is slow.

In terms of multistage interference cancellation we can view the gradient method of steepest descent by defining the negative gradient $\boldsymbol{\delta} = -\nabla\phi(\mathbf{z})$ and writing the decision statistic at stage k as

$$\begin{aligned}\mathbf{z}^{(k)} &= \mathbf{z}^{(k-1)} + \alpha_k\boldsymbol{\delta} \\ &= \mathbf{z}^{(k-1)} + \alpha_k\left(\tilde{\mathbf{y}} - \mathbf{R}\mathbf{z}^{(k-1)}\right) \\ &= \alpha_k\tilde{\mathbf{y}} - \left(\alpha_k\mathbf{R} - \mathbf{I}\right)\mathbf{z}^{(k-1)},\end{aligned} \tag{37}$$

[11] It should be noted that these two methods were investigated independently for the application of CDMA detection in [23].

where α_k is defined to minimize the error at each step, $\alpha_k = (\mathbf{z}^{(k-1)})^T\mathbf{z}^{(k-1)}/(\mathbf{z}^{(k-1)})^T\mathbf{R}\mathbf{z}^{(k-1)}$. Thus, this technique is similar to linear PIC as defined in (28) with the exception that a weighting factor α_k is used at each stage. Note that this method will converge provided $\mathbf{R}$ is not singular, although convergence time is directly related to the eigenvalue spread of $\mathbf{R}$.

Fig. 5 plots performance results using this method in the same system as the previous examples. It can be seen that the performance is identical to the decorrelator at low users for a relatively small number of stages (under 10). As the loading increases, the number of stages required to approach the decorrelator increases. An interesting phenomenon is that the performance at 10 stages is superior to the performance at later stages. We will address this in a moment.

A method that speeds up the convergence in gradient search techniques is to allow subsequent search directions to be non-orthogonal. Rather, subsequent directions are *conjugate*, that is they satisfy the relation $\mathbf{x}_i^H\mathbf{R}\mathbf{x}_j = 0$, $\forall\{i, j : i \neq j\}$. Thus, minimization in one direction does not interfere with minimization in previous directions. This allows convergence in K steps where K is the dimension of $\mathbf{R}$. Like the previous technique this can be thought of in terms of interference cancellation. The cancellation is similar to the steepest descent except that in addition to performing a weighted cancellation at each stage, a recursively estimated vector is also included in the cancellation.

The performance of the two gradient-search-based cancellation techniques is given in Fig. 6 as the number of stages is increased. The system is synchronous with an AWGN channel, E_b/N_o=8dB, 60 users, $N = 64$ random codes and random phases. It can be seen that the conjugate gradient based cancellation method converges within 60 stages, as expected while the steepest descent based algorithm requires 2000 stages to converge. An interesting result from Fig. 6 is that the gradient searches pass through "superior" solutions during convergence. That is the BER performance can be better before convergence.

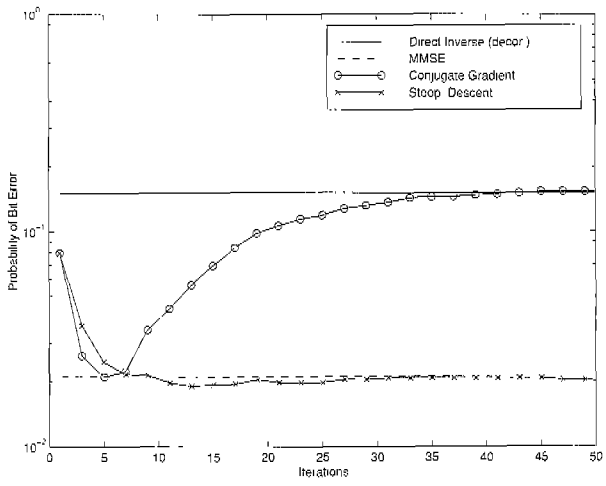

Fig. 7. Performance of "leaky" steepest descent algorithm and conjugate gradient algorithm vs. the number of stages of cancellation in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, 60 users, $\epsilon = 0.07$, random phase; direct inverse (decorrelator) and MMSE algorithms shown for comparison.)

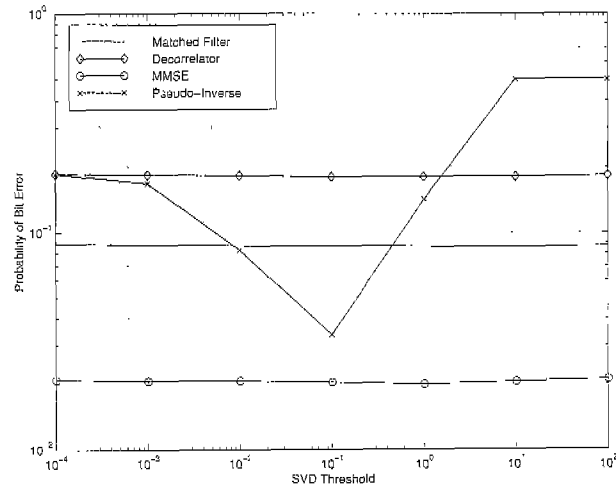

Fig. 8. Performance of pseudo-inverse vs. eigenvalue threshold in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, random phase; direct inverse (decorrelator) and MMSE algorithms shown for comparison.)

This is simply due to the fact that the decorrelator (direct matrix inversion) is not the optimal solution. In fact, at high loading factors the decorrelator provides performance that is significantly worse than the linear MMSE solution.

The performance of the gradient search algorithms can be understood as follows. The performance of the decorrelator is related to the decision statistic

$$\mathbf{z} = \mathbf{R}^{-1}\tilde{\mathbf{y}} = \mathbf{b} + \mathbf{R}^{-1}\mathbf{n}. \tag{38}$$

Thus, the values of the diagonal of $\mathbf{R}^{-1}$ are crucial as they will enhance the noise level, since they will be larger than unity. Further, these values will be directly related to the eigenvalues of $\mathbf{R}^{-1}$ which are the inverse of the eigenvalues of $\mathbf{R}$. To improve performance it is preferable not to include those components in the inversion which lead to large noise enhancement. Those components are the eigenvectors of $\mathbf{R}$ which have the smallest eigenvalues. However, the gradient search tends to find these modes last. The eigenvectors that have large eigenvalues will be captured first and relatively quickly. Thus if the search is stopped early before the other eigenvectors are found, performance is improved. In other words, if we limit the number of stages we improve the performance.

Rather than limiting the number of stages, another method to prevent convergence to the decorrelating solution is to add a "leak" to the gradient. That is,

$$\nabla\phi(\mathbf{z}) = \tilde{\mathbf{y}} - \mathbf{R}\mathbf{z} - \epsilon\mathbf{z}. \tag{39}$$

This results in the cancellation scheme defined by

$$\mathbf{z}^{(k)} = \alpha_k\tilde{\mathbf{y}} - \left(\alpha_k\mathbf{R} + (\epsilon\alpha_k - 1)\mathbf{I}\right)\mathbf{z}^{(k-1)}. \tag{40}$$

Note that the gradient can be rewritten as $\nabla\phi(\mathbf{z}) = \tilde{\mathbf{y}} - (\mathbf{R} + \epsilon\mathbf{I})\mathbf{z}$ showing that the algorithm will now converge to $\mathbf{z} = (\mathbf{R} + \epsilon\mathbf{I})^{-1}\tilde{\mathbf{y}}$. If $\epsilon = 0.5\sigma^2/a^2$ the algorithm converges to the linear MMSE solution. Fig. 7 shows the convergence of the steepest descent algorithm with a leak of $\epsilon = 0.07$ for the same system parameters as Fig. 6. As expected, the algorithm converges to the MMSE solution. Since the leak adds a constant to the eigenvalues of $\mathbf{R}$, it sets a floor on the minimum eigenvalue and prevents the large noise enhancement, improving performance. Another consequence is that convergence is relatively fast (within 10 stages).

B.4 Pseudo Inverse Technique

The previous section showed that if a leak were added to the convergence of the gradient search algorithm, the solution will not find the eigenvectors corresponding to the eigenvalues smaller than the leak size. Since these values are the source of the noise enhancement at large loading factors, performance is improved. Another method of accomplishing this task is to calculate a pseudo-inverse of $\mathbf{R}$ throwing away values from the singular value decomposition (SVD) below a certain threshold. Fig. 8 plots the performance of the pseudo-inverse as the SVD threshold is varied from 10^{-4} to 10^{2}, again for the same system parameters as Figs. 6 and 7. Note that by removing the eigenvectors corresponding to the smallest eigenvalues from the inverse (between 10^{-3} and 10^{-1}), the performance of the decorrelator can be improved by an order of magnitude.

B.5 Comparisons

A plot comparing each of the previously discussed linear cancellation techniques is presented in Fig. 9. It can be seen that the MMSE solution essentially underbounds[12] the performance of the linear cancellation techniques investigated. However, we see that the cancellation technique based on the gradient search can achieve the performance of the MMSE receiver. This performance requires either (1) the algorithm is stopped at the correct

[12]To within the accuracy of the simulated BER estimates. Additionally, the MMSE receiver is not necessarily the linear receiver which achieves minimum BER [24].

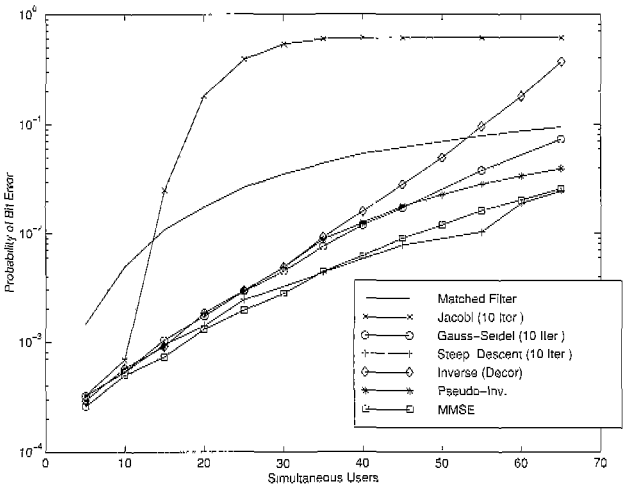

Fig. 9. Performance of several linear multiuser algorithms in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, random phase.)

stage or (2) a leak is included in the cancellation with the appropriate leak size. The knowledge necessary for the latter option is similar to the knowledge required for a direct MMSE solution, viz., the SNR. It appears that stopping the algorithm after 10 stages for steepest descent based or 5 stages for conjugate gradient based provides the best performance over all loadings considered. This will be directly related to the SNR which may be known in the average sense if a power control algorithm is employed.

It should be noted that each of the interference cancellation approaches investigated in this section do not require explicit channel information to perform cancellation. Phase knowledge is required to perform detection since coherent modulation is used, but no channel knowledge is used for cancellation. This can be most clearly seen from equations (27) and (33). This may seem somewhat counter-intuitive for a cancellation approach. However, this is obvious when one realizes that each of the "cancellation" approaches was trying to approximate the decorrelating receiver which requires only spreading code and timing information of the interferers (beyond what is required for the matched filter). Channel knowledge would be required for the MMSE receiver and is thus helpful for implementing the leak in the gradient-search-based approaches since the leak ideally should be related to the SNR.

C. Another View: The Bias Problem

A major result of the previous section is that linear parallel interference cancellation (i.e., PIC using only linear soft estimates), is equivalent to the Jacobi iteration method for solving $\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{y}$. This equivalence indicates that linear PIC will have poor performance at high loading factors since the Jacobi method will not converge if $\mathbf{R} - \mathbf{I}$ has an eigenvalue with absolute value greater than one. Another way of examining the performance of linear PIC is to examine the decision statistic. It can be shown that the decision statistic of linear PIC after cancellation is no longer unbiased. To show this we consider the phase corrected output of the matched filter for user 1

$$y_1 = \sqrt{P_1}b_1 + I_1 + n_1. \tag{41}$$

Assume for the sake of clarity a synchronous system such that $\mathcal{R} = \mathbf{R}(0) = \mathbf{R}$, I_1 and n_1 represent the interference and thermal noise seen in the matched filter output of user 1 and we ignore the dependence of y on the bit number. Note that y_1 is not independent of $\{b_k\}$, i.e., the data symbols of the interferers, due to the dependence of I_1 on $\{b_k\}$. Thus, when the set $\{y_k\}$, $2 \le k \le K$ is used as estimates of $\{a_k b_k\}$ to cancel interference from y_1, the expected value of y_1 after cancellation will be less than $\sqrt{P_1}b_1$ due to the dependence of each y_k on b_1. To demonstrate this we take the expected value of $y_1 - \hat{I}_1$ conditioned on b_1. First, recall the definition of $\hat{I}_1$ using $\hat{a}_k\hat{b}_k = y_k$ and using external phase estimates:

$$\begin{aligned}
\hat{I}_1 &= \sum_{k=2}^{K} y_k \cos(\phi_k - \phi_1)\,\rho_{1,k} \\
&= \sum_{k=2}^{K} \sqrt{P_k}b_k \cos(\phi_k - \phi_1)\,\rho_{1,k} \\
&\quad + \sum_{k=2}^{K} \left[\sum_{i\neq k} \sqrt{P_i}b_i\rho_{k,i}\cos(\phi_i - \phi_k)\right] \cos(\phi_k - \phi_1)\,\rho_{1,k} \\
&\quad + \sum_{k=2}^{K} n_k \cos(\phi_k - \phi_1)\,\rho_{1,k}.
\end{aligned} \tag{42}$$

Thus, the expected value of $y_1 - \hat{I}_1$ conditioned on b_1 is given by (43) as shown at the bottom of the next page, where for synchronous transmission $\rho_{k,1} = \rho_{1,k}$ and $\mathrm{var}(\rho_{1,k}) = \frac{1}{N}$ [25] and we have defined $\tilde{n}_1 = n_1 - \sum_{k=2}^{K} n_k \cos(\phi_k - \phi_1)\,\rho_{1,k}$. As the system loading grows, the decision statistic after one stage of cancellation becomes more biased toward the decision boundary resulting in decision errors if only one stage of cancellation is performed. If multiple stages of cancellation are used, this error will be compounded in future stages. This reduction in the mean of the decision statistic can be devastating to performance as will be seen. This insight provides the motivation for a final linear cancellation technique, partial PIC.

D. Partial Cancellation

One method of overcoming the problem of decision statistic bias in linear PIC is to perform partial interference cancellation [26], [27]. In partial interference cancellation, we multiply the estimate of the channel and symbol by a factor, ξ, less than unity. By attempting to cancel only part of the interference we introduce only part of the bias. By keeping the bias small at the first and second stages of cancellation, overall performance is improved. By the third stage, a partial cancellation factor is usually no longer needed since the bias becomes insignificant at this point.[13] A second benefit of this approach stems from the fact

[13]It can be shown that the bias is inversely proportional to the stage of cancellation.

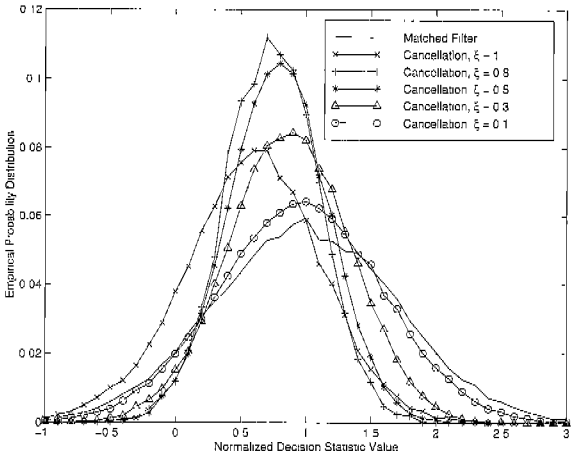

Fig. 10. Probability distribution function of decision statistic at matched filter and after one stage of cancellation using soft symbol estimates for different cancellation factors. ($\xi \in \{0.1, 0.3, 0.5, 0.8, 1.0\}$) (55 users, $N = 64$, synchronous AWGN channel,E_b/N_o = 8dB.)

that the reliability of the symbol decisions (and thus the channel estimate) is lower at initial stages of cancellation. When we consider that cancellation based upon an incorrect symbol estimate *adds* interference, partial cancellation mitigates this effect by reducing the penalty associated with incorrect symbol decisions. This can reduce the mean square error between the actual and estimated signals [8].

Partial cancellation is performed by multiplying the estimated signal by a factor less than unity prior to cancellation. That is

$$z_i^{(k)} = y_i - \sum_{j \neq i} \xi_k z_j^{(k-1)} \rho_{j,i}. \tag{44}$$

If we include a cancellation factor ξ_1 in the preceding derivation of (43) we find that the bias in the decision statistic after one stage of cancellation is

$$E\left[z_i^{(1)}\right] = \sqrt{P_i} b_i \left(1 - \xi_1 \frac{K-1}{2N}\right). \tag{45}$$

Fig. 10 plots the normalized (with respect to $\sqrt{P_i} b_i$) histogram of the decision statistic after one stage of cancellation when using linear soft symbol estimates with partial cancellation factors as in equation (44) with ξ_1=1 (full cancellation) and ξ_1=$\{0.8, 0.5, 0.3, 0.1\}$. Also plotted is the normalized histogram of the matched filter outputs (i.e., before cancellation). These histograms were taken from a simulation of 55 simultaneous, synchronous BPSK DS-CDMA signals operating in AWGN with $\frac{E_b}{N_o}$=8dB and a spreading gain of 64. Full cancellation results in a large bias in the statistic ($\overline{y_1 - \hat{I}_1} = 0.58$). Using a partial cancellation of 0.1 nearly eliminates the bias ($\overline{y_1 - \hat{I}_1} = 0.96$), but does not significantly reduce the variance as compared to stage 1, since little interference is removed. Using a factor of 0.8 both improves the mean ($\overline{y_1 - \hat{I}_1} \approx 0.66$) and reduces the variance when compared to full cancellation. A factor of $\xi_1 = 0.5$ appears to provide the best performance in this situation. This improved performance is then passed to later stages as demonstrated in the results of Fig. 11.

Fig. 11 shows the simulated performance of parallel cancellation for one, two and four stages of cancellation using a one symbol linear estimate with and without a partial cancellation factor. For the partial cancellation approach, the cancellation factor is ξ=0.4 in the first stage and ξ=0.8 at the second stage (further stages use $\xi = 1$). We can see that partial cancellation dramatically improves performance. There are several interesting points made by this figure. First, note that for high loading full cancellation, i.e., $\xi = 1$, performance is severely degraded, while partial cancellation degrades more gracefully. Moreover, the performance is degraded so much in the first stage that additional stages worsen system performance as predicted by our analysis in the last section. Finally, it should be noted that at a

$$\begin{aligned}
E\left[y_1 - \hat{I}_1 \middle| b_1\right] &= E\left[\sqrt{P_1} b_1 + \sum_{k=2}^{K} \sqrt{P_k} b_k \cos(\phi_k - \phi_1)\rho_{1,k} + n_1 - \sum_{k=2}^{K} \sqrt{P_k} b_k \cos(\phi_k - \phi_1)\rho_{1,k}\right. \\
&\quad \left. - \sum_{k=2}^{K} \sum_{i \neq k} \sqrt{P_i} b_i \cos(\phi_i - \phi_k)\cos(\phi_k - \phi_1)\rho_{k,i}\rho_{1,k} - \sum_{k=2}^{K} n_k \cos(\phi_k - \phi_1)\rho_{1,k} \middle| b_1\right] \\
&= E\left[\sqrt{P_1} b_1 - \sum_{k=2}^{K} \sum_{i \neq k} \sqrt{P_i} b_i \cos(\phi_i - \phi_k)\cos(\phi_k - \phi_1)\rho_{k,i}\rho_{1,k} + \tilde{n}_1 \middle| b_1\right] \\
&= E\left[\sqrt{P_1} b_1 - \sum_{k=2}^{K} \sqrt{P_k} b_1 \rho_{k,1}\rho_{1,k} \cos^2(\phi_k - \phi_1)\right. \\
&\quad \left. - \sum_{k=2}^{K} \sum_{\substack{i \neq k \\ i \neq 1}} \sqrt{P_i} b_i \cos(\phi_i - \phi_k)\cos(\phi_k - \phi_1)\rho_{k,i}\rho_{1,k} + \tilde{n}_1 \middle| b_1\right] \\
&= \sqrt{P_1} b_1 - \sqrt{P_k} b_1 \sum_{k=2}^{K} E\left[\rho_{1k}^2 \cos^2(\phi_k - \phi_1)\right] = \sqrt{P_1} b_1 \left(1 - \frac{K-1}{2N}\right)
\end{aligned} \tag{43}$$

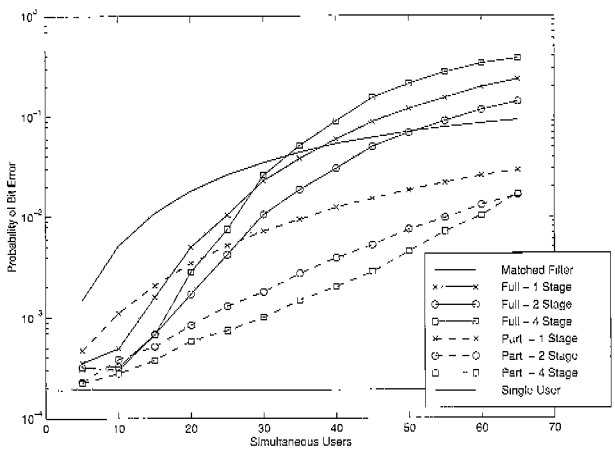

Fig. 11. Probability of bit error for full and partial linear parallel interference cancellation versus system loading. ($\xi_1 = 0.4$, $\xi_2 = 0.8$, $\xi_3 = 1$) using BPSK in Synchronous Channel and Soft Symbol Estimates ($E_b/N_o = 8$dB, $N = 64$.)

low number of users, partial cancellation performs worse than full cancellation after only one stage. This is due to the fact that the receiver is not cancelling as much interference as it could reliably. In other words, the partial cancellation ξ is too small for low loadings. Optimizing ξ to loading would improve performance. However, note even at this low number of users, the bias reduction in the first stage, causes performance in the second stage to be improved over full cancellation.

IV. NON-LINEAR CANCELLATION

In this section we investigate the performance of PIC when the linear constraint is relaxed. This allows non-linear symbol estimates to be used, but also makes an explicit channel estimation block necessary for cancellation. In section IV-A we first describe a non-linear MMSE[14] symbol estimator for linear modulation schemes such as BPSK. Then in section IV-B we consider the effect of channel estimation on performance. Section IV-C expands the results of IV-A for BPSK to non-linear modulation techniques. Section IV-D extends these results further to the far more common conditions of Rake reception of CDMA signals. Finally, section IV-E presents a few simulation results which lend support to the theoretical development presented.

A. Non-linear Symbol Estimation

The most obvious non-linear symbol estimate is to simply hard limit the matched filter output. This requires a separate channel estimation block, but will improve PIC performance provided the channel estimate is of reasonable quality. A better symbol estimate, however, is one that minimizes the effect of symbol estimation errors on system performance. It was shown previously that partial cancellation could significantly improve system performance, which suggests a search for an "optimal" symbol estimate.

[14]Note that unlike the MMSE estimator discussed in section III, this is a *non-linear* MMSE estimator.

In order to minimize the effect of incorrect interference estimation, we need to minimize a meaningful measure of the reconstruction error at the output of the reconstruction block. A natural and intuitive cost function is the mean-squared error (MSE). Consequently, the design problem for the reconstruction block in the first stage under the unconstrained (not constrained to be linear) MMSE criterion can be posed as follows: Given the received vector $\mathbf{r}$, i.e., a sampled version of (1), compute user k's received signal estimate $\hat{\mathbf{r}}_k$ that minimizes $E\left\{\|\mathbf{r}_k - \hat{\mathbf{r}}_k\|_2^2\right\}$, where $\|\mathbf{x}\|_2$ is the 2-norm of vector $\mathbf{x}$, and $E\{\cdot\}$ is the expectation operator.

The objective therefore is to compute

$$\hat{\mathbf{r}}_k = h_k(\mathbf{r}), \tag{46}$$

where

$$h_k = \arg\min_h E\left\{\|\mathbf{r}_k - h(\mathbf{r})\|_2^2\right\}. \tag{47}$$

The well-known solution to the non-linear MMSE estimation problem in the case of scalars is the conditional mean estimator [28]. Note the difference between the above equation and (10). Using a straight-forward extension of this result to the case of vector estimation, the non-linear MMSE estimator for user k's received signal is given by

$$\hat{\mathbf{r}}_k = h_k(\mathbf{r}) = E\{\mathbf{r}_k|\,\mathbf{r}\}. \tag{48}$$

It is shown in [8] and [12] that (48) evaluates to

$$\hat{\mathbf{r}}_k = \tilde{a}_k \hat{b}_k \mathbf{s}_k = \tilde{a}_k \tanh\left(\frac{\Re\left(\tilde{a}_k^* \tilde{y}_k\right)}{\sigma_k^2}\right)\mathbf{s}_k, \tag{49}$$

for BPSK modulation, where $\tilde{a}_k = \sqrt{P_k}e^{j\theta_k}$ is the complex channel induced upon the transmitted signal of user k[15], $\mathbf{s}_k$ is the spreading sequence, σ_k^2 is the variance of interference plus noise seen by channel k, $\tilde{y}_k$ is the matched filter output, and $\hat{b}_k$ is the MMSE symbol estimate.

This solution suggests the intuitively satisfying result that user k's signal should be reconstructed by passing the correlator output $\tilde{y}_k$ through the hyperbolic tangent nonlinearity, scale it with the complex amplitude $\tilde{a}_k$ and respread it with the spreading code $\mathbf{s}_k$. The hyperbolic tangent function increases monotonically from -1 at negative infinity to +1 at positive infinity. It may be observed that the optimal reconstruction block differs from that in the conventional non-linear PIC in that the hard-limiter is replaced by a sigmoidal nonlinearity. The sigmoidal nonlinearity endows the non-linear reconstruction block with the MMSE property.

We examine the performance of non-linear PIC using a hard-limit symbol estimate and the MMSE symbol estimate in Fig. 12. In this simulation a synchronous system is again assumed with random phases, an AWGN channel, random spreading codes, E_b/N_o=8dB, and N=64. Perfect channel knowledge is assumed. The performance of both the hard and soft non-linear symbol estimation approaches perform significantly

[15]Channel coefficient $\tilde{a}_k$ would, in lieu of genie assistance, have to be provided by a separate channel estimation block.

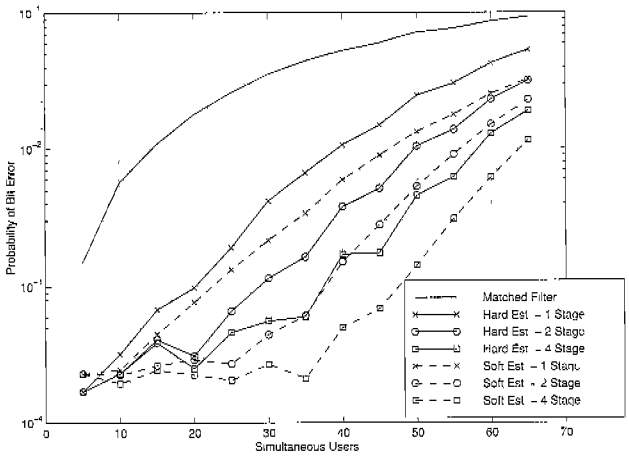

Fig. 12. Performance of non-linear parallel interference cancellation using soft and hard symbol estimates. (BPSK, AWGN, $E_b/N_o = 8$dB, $N = 64$, random codes, synchronous.)

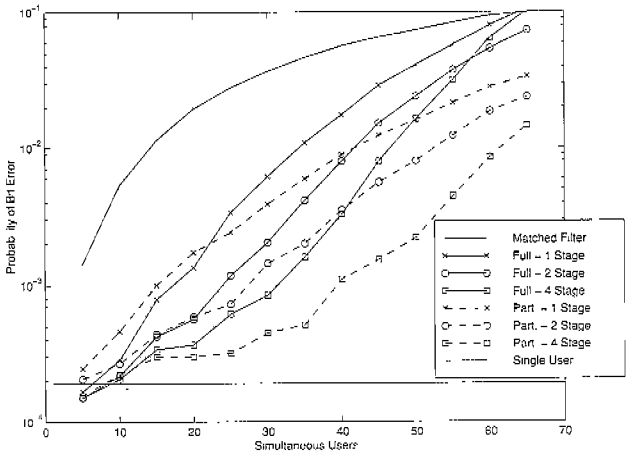

Fig. 13. Probability of bit error for full and partial PIC versus system loading ($\xi_1 = 0.4$, $\xi_2 = 0.8$,$\xi_3 = 1$) using BPSK in synchronous channel and hard symbol estimates with 6-symbol channel estimation. ($E_b/N_o = 8$dB, $N = 64$.)

better than the linear approach (without partial cancellation) as shown in Fig. 11. Additionally, we see that MMSE symbol estimation provides some performance advantage. At a BER of 10^{-2}, MMSE cancellation provides a 5-10% improvement in the number of supported users as compared to hard symbol estimates. Note that the extension to Rake reception is discussed in section IV-D.

B. Channel Estimation

By decoupling channel and symbol estimation, some method for explicit channel estimation is now clearly required to perform interference cancellation. The results of section IV-A assumed channel knowledge was provided by a genie. In section III it was shown that linear PIC produces a biased decision statistic after one stage of cancellation. It can be shown that bias also occurs in non-linear estimation when correlator outputs are used for channel estimation. This bias is directly related to the number of symbols used to estimate the channel. Thus, for an ν-symbol channel estimate (averaging or linear regression methods), the bias may be reduced by a factor of ν.

The performance of non-linear symbol estimation combined with channel estimation is examined in Figs. 13 through 15. The performance of hard-limit symbol estimates with channel estimates created using a six-symbol average are presented in Fig. 13. The system is synchronous, with an AWGN channel, $N = 64$, E_b/N_o=8dB and codes are random from symbol to symbol. Note that the performance of non-linear PIC with hard estimates and channel estimation degrades when compared to perfect channel knowledge as presented in Fig. 12. This degradation is due to both channel estimation error and decision statistic bias. The latter effect is evident by the performance of partial cancellation with six-symbol estimation also given in Fig. 13. We see that the performance of partial cancellation is improved significantly compared to full cancellation when channel estimation is based on correlator outputs. Again, the use of correlator outputs biases the decision statistic after cancellation. This bias can be eliminated by using outside channel

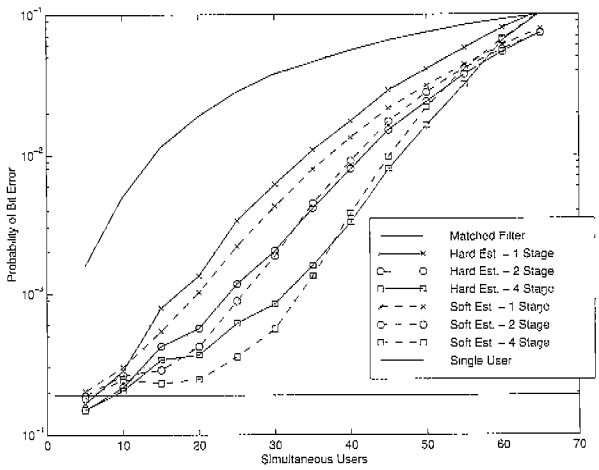

Fig. 14. Performance of non-linear parallel cancellation using soft and hard symbol estimates with 6-symbol channel estimates. (BPSK, AWGN, $E_b/N_o = 8$dB, $N = 64$, random codes, synchronous.)

estimates or mitigated by using partial cancellation. It should also be noted that in the absence of decision statistic bias, partial cancellation provides some benefit due to the reliability of the symbol estimates [29], [8].

The effect that channel estimation has on the MMSE symbol estimate is shown in Fig. 14. The system is synchronous, with an AWGN channel, $N = 64$, E_b/N_o=8dB and the codes are random from symbol to symbol. The MMSE symbol estimate was derived based on perfect knowledge of the channel. When the channel is estimated, the MMSE property no longer holds. Thus, the performance is degraded more than the hard-limit estimate as seen by comparing the performance with Fig. 12.

Finally, we examine the performance of non-linear successive interference cancellation. In Fig. 15 we present the performance of successive cancellation when hard estimates of the symbols are used along with six-symbol averages of the correlator outputs for channel estimation. Non-linear estimation provides significant improvement in the performance relative to the linear

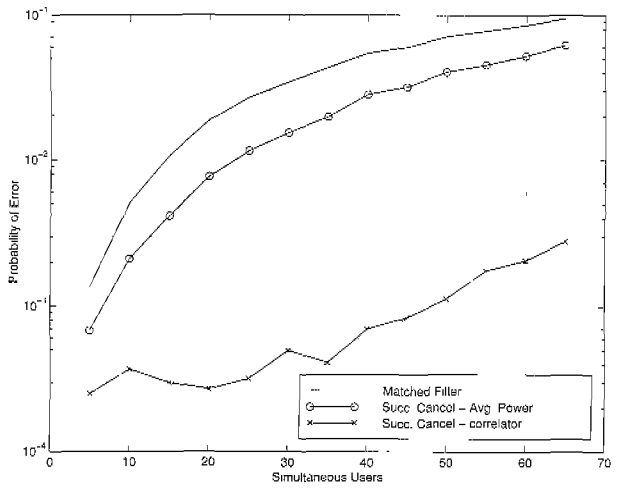

Fig. 15. Performance of successive interference cancellation using six symbol estimates for cancellation in AWGN channel. (synchronous system, random codes, spreading gain $N = 64$, $E_b/N_o = 8$dB, random phase; for Avg. Powers, ordering done on average power basis; for correlator, ordering done on basis of correlator outputs after each cancellation.)

estimation examined in Fig. 4, when ordering is based on reliability. Note that the performance is unchanged when average powers are the basis of ordering. This is because in an AWGN channel all users are received with equal average powers. Thus, the ordering criterion equates to random ordering. The reliability criterion requires significantly more computations since reliability must be re-evaluated after each cancellation, but provides far superior performance.

C. Non-linear M-ary Orthogonal Modulation

The previous discussion focused on linear modulation techniques, specifically BPSK. However, it is useful to explore the effect of non-linear modulation. With non-linear modulation, the linear multiuser detection techniques discussed earlier such as the decorrelating and linear MMSE receivers are not applicable. However, techniques such as non-linear PIC or SIC are appropriate. In this section we focus on the application of non-linear PIC to M-ary orthogonal modulation.

M-ary orthogonal modulation involves mapping a block of $\log_2 M$ bits on to one of M codewords, each M bits long, taking values from $\{-1, +1\}$ and orthogonal to each other. In CDMA systems, the resulting bit stream consisting of the M-ary symbols (or codewords) is then spread by multiplying with the spreading code and transmitted on a carrier. The set of M orthogonal codewords is usually chosen to be Walsh-Hadamard codes of length M. With such modulation, the received signal of the first user can be written as $\mathbf{r}_1 = \tilde{a}_1 \mathbf{x}_1 = \tilde{a}_1 \mathbf{S}_1 W \mathbf{i}_1$, where $\mathbf{S}_1$ is a diagonal matrix with the spreading code $\mathbf{s}_1$ on its diagonal, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M]$ is a matrix whose columns are the M Walsh codes, $\mathbf{i}_1$ is an indicator vector that picks the transmitted symbol from the Walsh matrix $\mathbf{W}$ and $\tilde{a}_1$ is the complex channel distortion. If the mth Walsh code is transmitted by user 1, then $\mathbf{i}_1 = \mathbf{e}_m$, where $\mathbf{e}_m$ is the mth column of the MxM identity matrix, so that $\mathbf{W}\mathbf{i}_1$ is the symbol that is transmitted. The received vector $\mathbf{r}$ can then be represented by

$$\mathbf{r} = \tilde{a}_1 \mathbf{S}_1 \mathbf{W} \mathbf{i}_1 + \mathbf{v}, \tag{50}$$

where $\mathbf{v} = \mathbf{r}_2 + \ldots + \mathbf{r}_K + \mathbf{n}_o$ is the interference plus noise component of the received signal. Let the number of samples in one Walsh symbol be N_s, so that the received signal $\mathbf{r}$ and the Walsh code vectors are all N_s-dimensional. Equation (48) can be used again to find the non-linear MMSE estimate of $\mathbf{r_1}$ as follows:

$$\hat{\mathbf{r}}_1 = E\{\mathbf{r}_1 | \mathbf{r}\} = \tilde{a}_1 \mathbf{S}_1 \mathbf{W} E\{\mathbf{i}_1 | \mathbf{r}\}. \tag{51}$$

The solution for this estimate is similar to that presented in section IV-A, only much more involved. It is shown in [12] to be

$$\hat{\mathbf{r}}_k = \tilde{a}_k \mathbf{S}_k \mathbf{W} \hat{\mathbf{i}}_k, \tag{52}$$

where $\hat{\mathbf{i}}_k = \left[\hat{i}_{k1}, \hat{i}_{k2}, \ldots, \hat{i}_{kM}\right]^T$, and

$$\hat{i}_{km} = \frac{\exp\{\Re(\tilde{a}_k^* \tilde{y}_{km})/\sigma_k^2\}}{\sum_{q=1}^{M} \exp\{\Re(\tilde{a}_k^* \tilde{y}_{kq})/\sigma_k^2\}}, \quad m = 1, 2, \ldots, M. \tag{53}$$

Equation (53) suggests the following steps in the optimal reconstruction of the kth user's received signal:

1. Despread the received signal to form $\mathbf{S}_k^H \mathbf{r}$
2. Compute the complex correlations of the despread signal with the M Walsh symbols to form the matched filter output $\tilde{\mathbf{y}}_k = \mathbf{W}^H(\mathbf{S}_k^H \mathbf{r})$. This can be efficiently computed using the Fast Hadamard Transform (FHT).
3. Pass the correlator outputs through sigmoidal nonlinearities given by (53) to obtain $\hat{\mathbf{i}}_k$, the vector of weights associated with each of the M symbol waveforms in the reconstructed symbol waveform.
4. Combine all the M symbol waveforms weighted by the corresponding elements in $\hat{\mathbf{i}}_k$ to obtain the reconstructed symbol waveform $\mathbf{W}\hat{\mathbf{i}}_k$. Observed that this amounts to computing the inverse FHT (IFHT) of $\mathbf{i}_k$. The IFHT, which is the same as the FHT within positive scaling, can therefore be used to efficiently compute the reconstructed symbol.
5. Spread the reconstructed symbol by multiplying with the spreading code and scale it by the complex channel amplitude $\tilde{a}_k$.

D. Extension to Rake Reception of Non-linear Modulation

The problem of optimal reconstruction for PIC with M-ary orthogonal modulation with Rake reception is now considered in this section. Let $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \ldots, \mathbf{r}^{(L)}$ be the received signals from L Rake fingers. The vectors are formed after individually synchronizing time for the L paths. The lth received signal is given by

$$\hat{\mathbf{r}}^{(l)} = \mathbf{r}_1^{(l)} + \mathbf{r}_2^{(l)} + \ldots + \mathbf{r}_K^{(l)} + \mathbf{n}^{(l)}, l = 1, 2, \ldots, L, \tag{54}$$

where $\mathbf{r}_k^{(l)}$ is the kth user's received signal and $\mathbf{n}^{(l)}$ is the noise on the lth finger. The problem now is to find the non-linear MMSE estimate of the kth user's received signal in the lth path,

given the received signals on L fingers of the Rake. In a manner similar to that presented in section IV-C, the solution may be found by considering

$$\hat{\mathbf{r}}_k^{(l)} = E\left\{ \mathbf{r}_k^{(l)} \middle| \tilde{\mathbf{y}}_k^{(1)}, \tilde{\mathbf{y}}_k^{(2)}, \ldots, \tilde{\mathbf{y}}_k^{(L)} \right\}, \tag{55}$$

where $\tilde{\mathbf{y}}_k^{(l)} = (\mathbf{S}_k\mathbf{W})^H \mathbf{r}_k^{(l)}$, $l = 1, 2, \ldots, L$, are the correlator outputs in the L finger of the Rake, each similar to $\tilde{\mathbf{y}}_k$ in section IV-C for a single path reception. The solution in (55) is now evaluated for the case of M-ary orthogonal modulation. With M-ary orthogonal modulation,

$$\mathbf{r}_k^{(l)} = \tilde{a}_k^{(l)} \mathbf{S}_k \mathbf{W} \mathbf{i}_k. \tag{56}$$

Denoting $\tilde{Y}_k = \left[\tilde{\mathbf{y}}_k^{(1)}, \tilde{\mathbf{y}}_k^{(2)}, \ldots, \tilde{\mathbf{y}}_k^{(L)}\right]$ and applying (56), (55) can be rewritten as

$$\hat{\mathbf{r}}_k^{(l)} = \tilde{a}_k^{(l)} \mathbf{S}_k \mathbf{W} E\left\{ \mathbf{i}_k | \tilde{Y}_k \right\}. \tag{57}$$

It follows that the optimal reconstructed signal can be expressed by

$$\hat{\mathbf{r}}_k^{(l)} = \tilde{a}_k^{(l)} \mathbf{S}_k \mathbf{W} \hat{\mathbf{i}}_k, \tag{58}$$

where $\hat{\mathbf{i}}_k = E\left\{ \mathbf{i}_k | \tilde{Y}_k \right\}$ is the estimate of $\mathbf{i}_k$. Note that, since $\hat{\mathbf{i}}_k$ is independent of l , the reconstructed symbol will be the same on all the fingers. The correlator outputs for the kth user are given by

$$\tilde{\mathbf{y}}_k^{(l)} = \tilde{a}_k^{(l)} \mathbf{i}_k + \mathbf{v}_k^{(l)}, l = 1, 2, \ldots, L. \tag{59}$$

As is the case in single path reception, note that the interference plus noise vector $\mathbf{v}_k^{(l)}$ is zero-mean Gaussian with independent components, each with a variance $\sigma_{k,l}^2$, and is independent of the transmitted symbols. Further, it may be assumed that the interference plus noise vectors in different fingers of the Rake are mutually independent. Since $\mathbf{i}_k$ can select any of the columns of the $M \times M$ identity matrix with equal probability, we have

$$\hat{\mathbf{i}}_k = E\left\{ \mathbf{i}_k | \tilde{Y}_k \right\} = \sum_{m=1}^{M} \mathbf{e}_m P[\mathbf{i}_k = \mathbf{e}_m | \tilde{Y}_k]. \tag{60}$$

The non-linear MMSE estimator for the kth user's received signal in the lth path, with M-ary orthogonal modulation and a Rake receiver with L fingers, is shown in [12] to be

$$\hat{\mathbf{r}}_k^{(l)} = \tilde{a}_k^{(l)} \mathbf{S}_k \mathbf{W} \hat{\mathbf{i}}_k, \tag{61}$$

where $\hat{\mathbf{i}}_k = \left[\hat{i}_{k1}, \hat{i}_{k2}, \ldots, \hat{i}_{kM}\right]^T$, and

$$\begin{aligned} \hat{i}_{km} &= \frac{\exp(z_m)}{\sum_{q=1}^{M} \exp(z_q)}, \quad m = 1, 2, \ldots, M, \\ z_m &= \sum_{l=1}^{L} \frac{\Re\left\{ \tilde{a}_k^{*(l)} \tilde{y}_{km}^{(l)} \right\}}{\sigma_{k,l}^2}, \quad m = 1, 2, \ldots, M. \end{aligned} \tag{62}$$

Concerning equations (61) and (62), we make the following observations:

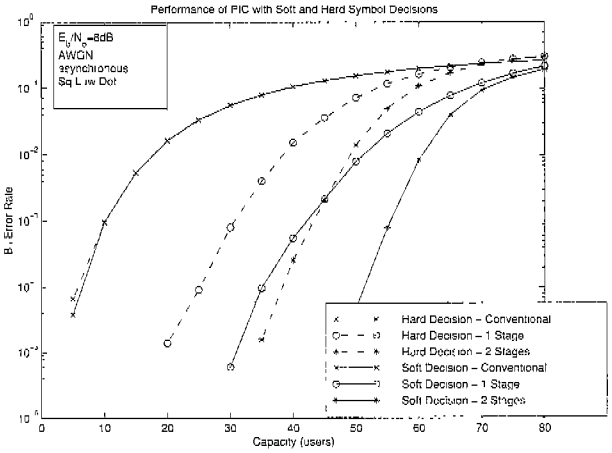

Fig. 16. Probability of bit error for parallel interference cancellation when 64-ary orthogonal modulation is used and hard or soft symbol estimation is employed.

1. The statistic z_m used in the computation of the MMSE reconstructed signal is simply the coherent maximal ratio combination (MRC) of L Rake fingers.
2. It follows that the correlator outputs on all Rake fingers should first be computed and then combined by MRC. The resulting statistic should then be passed through a sigmoidal non-linearity to obtain the reconstructed symbol. The reconstructed symbol waveform should then be appropriately delayed, spread and scaled by the complex amplitudes on all the fingers to obtain the multipath received signal estimates. Note that all the fingers use the same reconstructed symbol, though delayed and scaled differently.

E. Simulation Results

To verify the results given in preceding sections, PIC with both MMSE and hard-limit symbol estimation has been simulated for an asynchronous DS-CDMA system employing M-ary orthogonal modulation with $M = 64$. QPSK spreading was used with chips drawn from $\{e^{j\pi/4}, e^{j3\pi/4}, e^{-j\pi/4}, e^{-j3\pi/4}\}$. The number of PN chips per Walsh symbol was 256. Estimation of the complex channel amplitudes was conducted using a six-symbol average. The BER versus number of users is plotted in Fig. 16. Again, although the MMSE derivation assumed perfect channel knowledge, we have used a realizable channel estimate. However, even without perfect channel knowledge, it is clear that the MMSE symbol estimate provides large performance gains over hard-limiting.

V. CONCLUSIONS

In this work we have examined linear and non-linear interference cancellation techniques. The linear techniques are derived by exploring the relationship between PIC and the decorrelator (i.e., viewing PIC as a form of matrix inversion) and by examining the decision statistics after cancellation. It was shown that the linear techniques based on matrix inversion algorithms will converge to the decorrelator (with the exception of cancellation based on Jacobi iterations at moderate to high loading) for suffi-

cient stages. In high loading scenarios, however, the decorrelator performs significantly worse than the linear MMSE receiver. Linear interference cancellation techniques approach the linear MMSE solution prior to convergence. Thus, limiting the number of stages can improve performance. Some knowledge of the channel SNR is useful when determining the number of stages, similar to the knowledge needed for the linear MMSE receiver. Partial cancellation techniques based on decision statistic bias removal can also provide better performance than linear PIC without partial cancellation at lower computational cost. Additionally, it was shown that linear cancellation techniques can be applied which do not require phase/amplitude channel knowledge for cancellation. If non-coherent detection were employed, cancellation receivers can be employed which require only timing knowledge (as well as knowledge of the spreading codes).

Non-linear techniques were also investigated and the non-linear MMSE symbol estimate for BPSK as well as M-ary orthogonal modulation were presented. It was shown that the use of MMSE symbol estimates rather than a hard limit estimate can provide performance improvement especially in the case of M-ary orthogonal modulation. It was shown that partial cancellation is also useful in non-linear PIC due to the bias introduced through the channel estimate as well as symbol reliability. The gains of partial cancellation with hard-limited symbol estimates were similar to the gains seen by using MMSE symbol estimation for BPSK modulation. In general, non-linear techniques improve performance at the cost of higher computational complexity as well as a requirement for explicit channel estimation.

We conclude that linear techniques are likely to be most useful when computational limitations are more strict or when non-coherent detection is used. Since the linear techniques do not require explicit channel knowledge (outside of timing) they are more appropriate when channel estimation is not practical. Non-linear cancellation will provide superior BER performance but will require higher computation and explicit channel estimation. Additionally, non-linear techniques are applicable to a larger class of modulation schemes, including non-linear modulation. Also, partial cancellation can provide significant benefits in both linear and non-linear PIC with little computational burden.

ACKNOWLEDGMENTS

The authors would like to acknowledge their colleagues within Lucent Technologies' research community who contributed to this research: Howard Huang, Shang-Chieh Liu, Shirish Nagaraj, Larry Ozarow, Dirck Uptegrove, and Michaela Vanderveen. Neiyer Correal and Brian Woerner of Virginia Tech's Mobile and Portable Radio Research Group also provided valuable discussions. The authors also acknowledge the input of Yih-Fang Huang of Notre Dame.

REFERENCES

[1] S. Verdú, "Minimum probability of error for asynchronous gaussian multiple access channel," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 1, pp. 85–96, Jan. 1986.

[2] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Personal Commun.*, vol. 2, no. 2, pp. 46–58, Apr. 1995.

[3] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, vol. 34, no. 10, pp. 124–137, Oct. 1996.

[4] S. Verdú, *Multiuser Detection*, Cambridge University Press, New York, 1998.

[5] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple access communications," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 509–519, Apr. 1990.

[6] R. Kohno, H. Imai, M. Hatori, and S. Pasupathy, "An adaptive canceller of cochannel interference for spread-spectrum multiple-access communication networks in a power line," *IEEE J. Select. Areas Commun.*, vol. 8, no. 4, pp. 691–699, May 1990.

[7] R. M. Buehrer and B. D. Woerner, "Analysis of multistage interference cancellation for CDMA using an improved Gaussian approximation," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1308–1316, Oct. 1996.

[8] D. Divsalar, M. K. Simon, and D. Raphelli, "Improved parallel interference cancellation," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 258–268, Feb. 1998.

[9] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform Theory*, vol. 35, no. 1, pp. 123–136, Jan. 1989.

[10] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, no. 4, pp. 683–690, May 1990.

[11] 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, no. 4, pp. 641–649, May 1990.

[12] S. Gollamudi, R. M. Buehrer, S. Nagaraj, and Y.-F. Huang, "Optimal multistage interference cancellation for CDMA systems using the nonlinear MMSE criterion," in *Proc. of the 32nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1998.

[13] R. Lupas and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 496–508, Apr. 1990.

[14] M. L. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 944–960, July 1995.

[15] S. Verdú, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Commun.*, vol. COM-39, no. 9, pp. 890–897, Sept. 1986.

[16] Z. Xie, C. K. Rushforth, and R. T. Short, "Multiuser signal detection using sequential decoding," *IEEE Trans. Commun.*, vol. 38, no. 5, pp. 578–583, May 1990.

[17] V. Ghazi-Moghadam, L. B. Nelson, and M. Kaveh, "Parallel interference cancellation for CDMA systems," in *Proc. of the 33rd Annual Allerton Conference*, Urbana-Champaign, IL, 1995.

[18] R. M. Buehrer and B. D. Woerner, "The asymptotic multiuser efficiency of M-stage interference cancellation receivers," in *Proc. of the Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 570–574, Helsinki, Finland, Sept. 1997.

[19] L. K. Rasmussen, D. Guo, T. J. Lim, and Yao Ma, "Aspects on linear parallel interference cancellation in CDMA," in *Proc. of the 1998 International Symposium on Information Theory*, p. 37, Cambridge, MA, 1998.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins, Third edition, 1996.

[21] 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, no. 5, pp. 796–807, June 1994.

[22] A. C. K. Soong and W. A. Krzymien, "A novel CDMA multiuser interference cancellation receiver with reference symbol aided estimation of channel parameters," *IEEE J. Select. Areas Commun.*, vol. 14, no. 8, pp. 1536–1547, Oct. 1996.

[23] M. J. Juntti, B. Aazhang, and J. O. Lilleberg, "Iterative implementation of linear multiuser detection for dynamic asynchronous cdma systems," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 503–508, Apr. 1998.

[24] X. Wang, W.-S. Lu, and A. Antoniou, "Constrained minimum-BER multiuser detection," in *Proc. of the 1999 International Conference on Acoustics, Speech and Signal Processing*, Pheonix, AZ, Mar. 15-19 1999.

[25] M. B. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication - part I: System analysis," *IEEE Trans. Commun.*, vol. COM-25, no. 8, pp. 795–799, Aug. 1977

[26] N. Correal, R. M. Buehrer, and B. D. Woerner, "A DSP-based DS-CDMA multiuser receiver based on partial parallel interference cancellation," *IEEE J. Select. Areas Commun* , vol. 17, no. 4, pp. 613–630, Apr. 1999.

[27] R. M. Buehrer, *The Application of Multiuser Detection to Cellular CDMA*, PhD thesis, Virginia Polytechnic Institute and State University, 1996.

[28] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1984.

[29] R. M. Buehrer and S. P. Nicoloso, "Comments on partial interference cancellation," *IEEE Trans. Commun* , vol. 47, no. 5, pp. 658–661, May 1999.

R. Michael Buehrer was born in Toledo, OH on February 7, 1969. He received his B.S.E.E. and M.S.E.E. form The University of Toledo in 1991 and 1993 respectively where he specialized in satellite communications. During 1994 and 1995 he worked as a systems engineer at Stanford Telecommunications in the Advanced Programs department in Reston, VA. In 1996 he was awarded a Ph. D. from Virginia Polytechnic Institute and State University where he studied under the Bradley Fellowship. While at Virginia Tech, Dr. Buehrer worked with the Mobile and Portable Radio Research Group (MPRG) in the areas of multiuser detection, spread spectrum applied to CDMA and mobile radio. In particular he specialized in interference cancellation techniques for cellular CDMA systems.

In August 1996, Dr. Buehrer joined Bell Laboratories' Wireless Communication Lab as a Member of Technical Staff where he has worked on advanced modulation and access techniques for ISM band communication. He is currently a member of the Wireless Signal Processing Group and is working on interference cancellation, adaptive antenna algorithms and downlink diversity methods for 2nd and 3rd generation CDMA and TDMA cellular systems.

Steven P. Nicoloso was born in Portland, OR on September 20, 1966. He received his B.S. from Liberty University, Lynchburg VA in 1988. From 1995 to 1997 he was a student researcher at the Mobile and Portable Radio Research Group at Virginia Tech, Blacksburg VA, where he worked on applying time-domain and space-domain signal processing to CDMA transceiver design. He received his M.S.E.E. from Virginia Tech in 1997. In June 1997, Mr. Nicoloso joined Lucent Technologies' Wireless Technology Laboratory in Whippany, NJ where he currently works on design, simulation, and implementation of forward-looking signal processing technologies in 2nd and 3rd generation CDMA transceivers.

Sridhar Gollamudi received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay in May, 1994, and the M.S.E.E. degree from University of Notre Dame in January, 1994. He was awarded the *Center for Applied Mathematics Fellowship* at the University of Notre Dame for the academic year 1996-97. From May to December 1997, he was an intern at Lucent Technologies-Bell Laboratories, Whippany, New Jersey, where he worked on interference suppression for wireless CDMA systems. He is currently a doctoral candidate at the Department of Electrical Engineering, University of Notre Dame.

Mr. Gollamudi is a student member of the IEEE and a member of Eta Kappa Nu. His research interests are in the areas of statistical and adaptive signal processing, communication theory, multiuser communication and antenna array processing.