Steady-State Analysis of the Adaptive Successive Interference Canceler for DS/CDMA Signals

Kuei-Chiang Lai and John J. Shynk, *Senior Member, IEEE*

***Abstract*—The adaptive successive interference canceler (ASIC) is a multistage receiver that sequentially detects and removes cochannel users from the received signal impinging on a single antenna element. Each stage of the ASIC consists of a conventional matched filter (MF) detector and an adaptive interference canceler (AIC) that employs the least-mean-square (LMS) algorithm to recursively estimate the received amplitude of the detected signal. In this paper, we investigate the performance of the ASIC using a Wiener model of convergence for the LMS algorithm, deriving expressions for the asymptotic mean and variance of the amplitude estimate and the steady-state bit error rate (BER). The analyses and computer simulations demonstrate that the performance of the ASIC exceeds that of the conventional SIC (CSIC), which utilizes the MF output as the received amplitude estimate.**

***Index Terms*—Adaptive signal detection, cochannel interference, code division multiple access, communication systems, land mobile radio cellular systems, least mean square methods.**

I. Introduction

DIRECT-SEQUENCE code-division multiple-access (DS/CDMA) techniques [1] have been receiving considerable attention in cellular mobile radio and personal communications services (PCS) because of their ability to mitigate multipath interference and the potential increase in the capacity of the communication system. However, multiple access interference (MAI) arising from the nonorthogonality between the spreading sequences is a significant limiting factor in the performance of such systems. It is well known that the conventional matched filter (MF) detector [2] suffers from the near–far problem; the detection of weaker users is severely degraded by the large MAI caused by stronger users. The optimum multiuser detector based on maximum-likelihood sequence detection was proposed in [3]. It has been shown that a huge performance gain can be achieved over the MF detector, and the near–far problem can be alleviated by exploiting the structure of the MAI to jointly estimate the user data. However, since the computational complexity of the optimum multiuser detector is prohibitively high for use in practical situations, research efforts have focused on suboptimum multiuser detectors (e.g., see [4] for a review of several linear and nonlinear techniques). The suboptimum approaches can achieve superior performance compared with the MF detector with a reasonable level of computational complexity.

Among the many multiuser detectors, interference cancellation (IC) techniques are relatively computationally efficient and can be used in systems utilizing aperiodic spreading sequences (i.e., spreading sequences with periods longer than the symbol interval) without incurring additional complexity. Basically, there are two IC architectures, both of which have a multistage structure: the parallel interference canceler (PIC) (e.g., see [5]) and the successive interference canceler (SIC) (e.g., see [6] and [7]). In each stage, the PIC *simultaneously* regenerates and cancels from each user the MAI due to other users based on the detected symbols in the preceding stage. The SIC *sequentially* removes the MAI due to the stronger users before detecting the weaker ones (i.e., only one user is detected and canceled in a stage); as a result, the near–far problem is less pronounced in the SIC than in the conventional MF detector. It was demonstrated via computer simulations that the SIC outperforms one- and two-stage PICs (all using linear cancellation) in fading channels [8].

In order to regenerate and cancel the MAI, accurate parameter estimation is necessary. The conventional SIC (CSIC) (which is also referred to as the linear SIC) [7] employs the magnitude of the output of the MF bank as the received amplitude estimate. An adaptive multiuser decorrelator, which minimizes a least-squares cost function of the received signal and its estimate, in combination with an IC structure, was proposed in [9] to jointly estimate BPSK-modulated data and the user amplitudes. However, this method cannot be applied in a straightforward manner to systems with higher dimensional modulation formats in fading channels. An SIC structure with a minimum mean-square-error (MMSE) receiver and a linear interference canceler embedded in each stage was introduced in [10]. Analyses of the bit error rate (BER) and the asymptotic multiuser efficiency (AME) based on a Gaussian approximation were presented for additive white Gaussian noise (AWGN) and flat-fading channels. More recently, decision-driven amplitude estimates have been used in conjunction with the PIC architecture. For example, in [11], each of the K active users is associated with a $(K-1)$-coefficient adaptive interference canceler (AIC) that attempts to remove the MAI from the other users without prior knowledge of their amplitudes. Using tentative decisions of the interfering users (thus, it is "decision-driven"), the AIC is updated at the symbol rate using a gradient algorithm to minimize the output power. A combination of this structure (but with canceler weights updated by a bootstrap algorithm) and a blind technique that suppresses unknown interference (e.g., intercell

Manuscript received August 17, 2000; revised June 29, 2001. This work was supported by Applied Signal Technology, Inc., and the University of California MICRO Program. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michail K. Tsatsanis.

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: shynk@ece.ucsb.edu).

Publisher Item Identifier S 1053-587X(01)08445-8.

interference) was described in [12]. Another adaptive PIC structure was proposed in [13], where a single K-coefficient AIC is adapted at the chip rate by the least-mean-square (LMS) algorithm [14] to jointly estimate the received amplitudes of the active users (i.e., each coefficient serves as the amplitude estimate for one user).

In this paper, we analyze the performance of an adaptive SIC (ASIC) structure that utilizes a single-tap AIC (similar to that in [13]) in each stage to recursively estimate and cancel the MAI of the detected users. We focus on the asymptotic mean and variance of the AIC weight (amplitude estimate) and the steady-state BER of the ASIC and compare the results with those of the CSIC via analysis and computer simulations. Although we consider a DS/CDMA system using periodic spreading sequences and BPSK modulation in an AWGN channel, the ASIC structure can be easily applied, with slight modifications, to practical systems with aperiodic spreading sequences and more complicated modulation formats in fading multipath channels. Specifically, multiple-tap AICs can be used to track multipath channels and to regenerate/cancel the MAI and interchip interference. See [15] and [16], where variants of the basic ASIC structure considered in this paper are used to demodulate the uplink and downlink data, respectively, of an Interim Standard 95 (IS-95) system [17]. The receiver structure analyzed here differs from that in [10] in two basic components: the amplitude estimator (decision-driven versus linear interference cancellation) and the receiver front end. It is interesting to note that exact expressions of the BER and the AME, as well as the mean and variance of the amplitude estimate used in [10], can be derived (without resorting to a Gaussian approximation) using the matrix approach in [18] and the results in Section V of this paper.

The rest of this paper is organized as follows. Section II introduces the signal model used in the analyses. The ASIC structure is described in Section III, and its steady-state properties are examined in Section IV. The corresponding properties of the CSIC are provided in Section V for comparison. Numerical examples and computer simulations of the ASIC and CSIC are presented in Section VI to illustrate their performance. Finally, conclusions are outlined in Section VII.

II. Signal Model

Consider a DS/CDMA system with K active users that transmit their information asynchronously over a common AWGN channel. The received signal at the base station can be modeled as

$$r(t)=\sum_{j=1}^{K}\sum_{i=-\infty}^{\infty}A_j b_j(i)s_j(t-iT-\tau_j)+n(t) \qquad (1)$$

where

- A_j received amplitude of the jth user;
- $b_j(i)$ transmitted binary symbol (± 1) of the jth user for $t\in[iT,(i+1)T]$;
- $s_j(t)$ signature waveform of the jth user;
- T symbol (bit) interval;
- τ_j time delay of the jth user;
- $n(t)$ AWGN with zero mean and two-sided power spectral density of σ^2 W/Hz.

The following additional properties are assumed. The signature waveforms are time limited in $[0,T]$, and each has unit energy, i.e., $s_j(t)=0$ for $t\notin[0,T]$ and $\int_0^T s_j^2(t)\,dt=1\ \forall j$. A signature waveform can be represented by

$$s_j(t)=\sum_{l=0}^{N-1}s_{j,l}P_{T_c}(t-lT_c) \qquad (2)$$

where N is the processing gain (number of chips per symbol), $\{s_{j,0},\ldots,s_{j,N-1}\}$ is the normalized spreading sequence (with values $\pm 1/\sqrt{N}$) assigned to the jth user, and

$$P_{T_c}(t)=\begin{cases}\dfrac{1}{\sqrt{T_c}}, & t\in[0,T_c]\\ 0, & \text{otherwise}\end{cases} \qquad (3)$$

is the rectangular chip waveform of duration T_c (note that $T_c=T/N$). The transmitted data $\{b_j(i)\}$ are independent and identically distributed (i.i.d.) $\forall\, i,\ j$; they take on the values ± 1 with equal probability. Without loss of generality, the users are labeled according to their signal strength, i.e., $A_1\geq A_2\geq\cdots\geq A_K$.

In order to keep the illustration simple, we restrict our attention to a synchronous system (i.e., $\tau_1=\cdots=\tau_K=0$); however, it is straightforward to extend the ASIC described below to an asynchronous system. Using a chip MF and sampling at the chip rate, the continuous-time received signal in (1) during the ith symbol interval $iT\leq t\leq(i+1)T$ can be converted to the following discrete-time N-vector

$$\mathbf{r}(i)=\sum_{j=1}^{K}A_j b_j(i)\mathbf{s}_j+\mathbf{n}(i) \qquad (4)$$

where

$$\begin{aligned}\mathbf{r}(i)&=[r_0(i),\ldots,r_{N-1}(i)]^T\\ r_l(i)&=\int_{iT+lT_c}^{iT+(l+1)T_c}r(t)P_{T_c}(t-iT-lT_c)\,dt\\ &\qquad l=0,1,\ldots,N-1\\ \mathbf{s}_j&=[s_{j,0},\ldots,s_{j,N-1}]^T\\ \mathbf{n}(i)&=[n_0(i),\ldots,n_{N-1}(i)]^T\end{aligned}$$

and

$$\begin{aligned}n_l(i)&=\int_{iT+lT_c}^{iT+(l+1)T_c}n(t)P_{T_c}(t-iT-lT_c)\,dt\\ &\qquad l=0,1,\ldots,N-1.\end{aligned}$$

It can be shown that $n_l(i)$ is a zero-mean Gaussian random variable with variance σ^2 and that $E[n_{l_1}(i_1)n_{l_2}(i_2)]=0$ for $l_1\neq l_2$ or $i_1\neq i_2$. The noise component at the output of the jth user's (symbol) MF is defined as $\hat{n}_j(i)\triangleq\mathbf{s}_j^T\mathbf{n}(i)$. It is straightforward to show that $\hat{n}_j(i)$ is a zero-mean Gaussian random variable with variance σ^2 and that $E[\hat{n}_j(i)\hat{n}_l(i)]=\rho_{jl}\sigma^2$, where $\rho_{jl}\triangleq\mathbf{s}_j^T\mathbf{s}_l$ is the cross-correlation between the spreading sequences of users j and l.

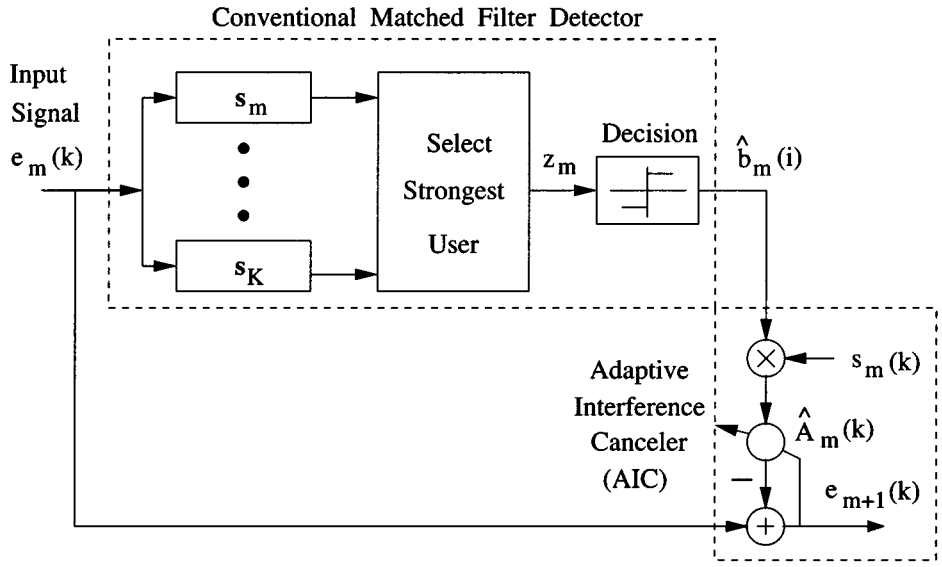

Fig. 1. Stage m of the ASIC consisting of a conventional MF detector and an adaptive interference canceler (AIC).

III. Adaptive Successive Interference Canceler (ASIC)

The ASIC detector is a multistage structure that sequentially recovers several users: one user in each stage. This system is reminiscent of the multistage constant modulus (CM) array [19], except that it employs a single antenna element (instead of an array/beamformer), and the CM criterion is not used for the SIC. Each stage of the ASIC consists of a conventional MF detector followed by an AIC that is adjusted by the LMS algorithm. An adaptive algorithm is employed so that the canceler weight can track time-varying channel variations in the wireless communication environment.

A block diagram of the mth stage of the ASIC is shown in Fig. 1. The conventional detector is essentially a MF bank followed by a hard decision device that detects only the strongest of the users remaining at that stage. This can be determined by choosing the user with the largest magnitude at the output of the MF bank. Without loss of generality, we assume that the mth user is detected in the mth stage. The conventional detector output of the mth stage in the ith symbol interval is given by

$$\hat{b}_m(i) = \operatorname{sgn}\left[\mathbf{s}_m^T \mathbf{e}_m(i)\right] \tag{5}$$

where

- sgn[·] signum function;
- $\mathbf{s}_m$ mth user's spreading sequence;
- $\mathbf{e}_m(i)$ input signal vector of the mth stage [note that $\mathbf{e}_1(i) = \mathbf{r}(i)$ for the first stage].

Specifically, $\mathbf{e}_m(i) = [e_m(k), \ldots, e_m(k+N-1)]^T$, where i and k are the symbol and chip indexes, respectively, which satisfy $k = iN$. The AIC respreads the detected data symbol $\hat{b}_m(i)$ using the spreading sequence $\mathbf{s}_m$ and the received amplitude estimate $\hat{A}_m(k)$ and then subtracts the respread signal from the input of that stage. The difference signal, which is given by

$$e_{m+1}(k) = e_m(k) - \hat{A}_m(k)\hat{b}_m(i)s_m(k) \tag{6}$$

becomes the input to stage $m+1$, where $s_m(k) \triangleq s_{m,(k)_N}$ is the corresponding chip in the mth user's spreading sequence [the subscript $(k)_N$ denotes k modulo N]. If the detected data symbol and the received amplitude estimate are both correct, then the next stage is free from the MAI caused by the detected user.

The LMS algorithm adapts the canceler weight as follows:

$$\hat{A}_m(k+1) = \hat{A}_m(k) + 2\mu e_{m+1}(k)\hat{b}_m(i)s_m(k) \tag{7}$$

where $\mu > 0$ is the step-size parameter. It should be noted that the subtraction in (6) and the canceler weight update in (7) are performed at the chip rate, which allows the canceler weight to track fast channel variations (fading). However, in a slow-fading environment, block adaptation (using, for example, a block LMS algorithm [20]) with a block size less than the channel coherence time [1] can be used to reduce the update rate and complexity of the AIC.

In order to motivate the canceler structure, we define the output power of the mth stage for a fixed $\hat{A}_m$ as

$$\begin{aligned}\xi^{(m)} &\triangleq \frac{1}{N} E\left[\mathbf{e}_{m+1}^T(i)\mathbf{e}_{m+1}(i)\right]\Big|_{\hat{A}_m(k)=\hat{A}_m} \\ &= \frac{1}{N} E\left[\sum_{k=iN}^{(i+1)N-1} \left(e_m(k) - \hat{A}_m\hat{b}_m(i)s_m(k)\right)^2\right]\end{aligned} \tag{8}$$

which is the output power averaged over one symbol interval (N chips). It is clear that if the decisions up to and including the mth stage are correct [i.e., $\hat{b}_j(i) = b_j(i)\ \forall\, i$ and $j = 1, \ldots, m$], then minimization of $\xi^{(m)}$ yields $\hat{A}_m = A_m$ (the actual received amplitude of the detected user) because the $\{b_j(i)\}$ are mutually uncorrelated. Based on this observation, the canceler weight of the mth stage is chosen to minimize the output power of that stage, i.e.,

$$\hat{A}_{m,o} = \arg\min_{\hat{A}_m} \xi^{(m)} \tag{9}$$

where the subscript o denotes the global minimum. Although, in general, the decision on $\hat{b}_m(i)$ is not always correct, it will be

demonstrated in the next section that the canceler weight thus obtained [i.e., the Wiener solution in (9)] is close to the actual received amplitude, provided the decisions on the data symbols are sufficiently reliable. Stochastic gradient algorithms such as the LMS algorithm in (7) can be used to approximate the solution in (9).

We should also mention that the ASIC reduces to the cascaded conventional MF detector when $\hat{A}_m(k) = 0\ \forall m, k$ and to the CSIC structure in [7] when the magnitude of the MF output is used as the canceler weight [i.e., $\hat{A}_m(k) = |\mathbf{s}_m^T \mathbf{e}_m(i)|$]. In addition, the conventional MF detector embedded in each stage can be replaced with other receivers, such as a decorrelator [21] or an MMSE receiver [22], to improve the performance (for symbol and decision-driven amplitude estimates) at the expense of an increase in the computational complexity (and limited applicability to systems using aperiodic spreading sequences). In the next section, we focus on systems that use a MF in the front end.[1]

IV. Steady-State Properties of the ASIC

As discussed previously, the AIC weight is an estimate of the received amplitude; its accuracy influences the performance of the ASIC. In this section, we examine the steady-state performance of the ASIC in terms of the asymptotic mean and variance of the canceler weight and the output power at convergence [23]. Using a Wiener model, if the step-size parameter is bounded by $0 < \mu < 1/E[(\hat{b}_m(i)s_m(k))^2] = N$ [14], the AIC weight in the mth stage converges in the mean to

$$\hat{A}_{m,o} = E\Big[\hat{b}_m(i)\mathbf{s}_m^T\mathbf{e}_m(i)\Big] \tag{10}$$

which is obtained by solving (9) using (8). The corresponding steady-state BER performance is also analyzed [24]. For notational convenience, the symbol index i is suppressed in the following derivations.

A. Asymptotic Mean of the AIC Weight for $m = 1$

In Appendix A, it is shown that the converged weight for the first stage is

$$\begin{aligned}\hat{A}_{1,o} &= A_1 E\Big[b_1\hat{b}_1\Big] + \sum_{j=2}^{K} A_j\rho_{1j}E\Big[b_j\hat{b}_1\Big] + E\Big[\hat{n}_1\hat{b}_1\Big] \\ &= A_1\Big(1 - 2P_e^{(1)}\Big) + 2^{-(K-1)}\sum_{j=2}^{K} A_j\rho_{1j} \\ &\quad \cdot \sum_{b_1}\cdots\sum_{b_{j-1}}\sum_{b_{j+1}}\cdots\sum_{b_K} \\ &\quad \cdot \left[Q\left(\frac{-A_j\rho_{1j} + \sum_{l\neq j} A_l b_l\rho_{1l}}{\sigma}\right)\right. \\ &\quad \left. - Q\left(\frac{A_j\rho_{1j} + \sum_{l\neq j} A_l b_l\rho_{1l}}{\sigma}\right)\right] \\ &\quad + 2^{-K}\sqrt{\frac{2\sigma^2}{\pi}}\sum_{b_1}\cdots\sum_{b_K}\exp\left(\frac{-\left(\sum_{l=1}^{K} A_l b_l\rho_{1l}\right)^2}{2\sigma^2}\right)\end{aligned} \tag{11}$$

where

$$P_e^{(1)} = 2^{-(K-1)}\sum_{b_2}\cdots\sum_{b_K} Q\left(\frac{A_1 + \sum_{l=2}^{K} A_l b_l\rho_{1l}}{\sigma}\right) \tag{12}$$

is the error probability of the first stage (MF stage), and $Q(x) \triangleq 1/\sqrt{2\pi}\int_x^\infty e^{-t^2/2}\,dt$ [1].

From this result, we conclude the following.

1) When no errors occur [i.e., $P_e^{(1)} = 0$ or, equivalently, $\sigma = 0$ and $A_1 > \sum_{l=2}^{K} A_l|\rho_{1l}|$], the second and third terms in (11) are zero because $\hat{b}_1 = b_1$ is uncorrelated with $\{b_j\}_{j=2}^K$ and $\hat{n}_1$. In this case, $\lim_{k\to\infty} E[\hat{A}_1(k)] = \hat{A}_{1,o} = A_1$, and the AIC weight converges in the mean to the actual received amplitude (i.e., it is an unbiased estimate).
2) When $P_e^{(1)} \neq 0$ and $\rho_{1j} \neq 0\ \forall j$, $\hat{b}_1$ is correlated with $\{b_j\}_{j=1}^K$ and $\hat{n}_1$, which implies that b_1 will not be completely removed by the AIC. Furthermore, $\{b_j\}_{j=2}^K$ will undergo partial cancellation in the first stage. In this case, $\hat{A}_{1,o}$ is a linear combination of $\{A_j\}_{j=1}^K$ and the noise term. Thus, the AIC weight is, in general, a biased estimate of the received amplitude (i.e., $\lim_{k\to\infty} E[\hat{A}_1(k)] = \hat{A}_{1,o} \neq A_1$), resulting in an amplitude mismatch. The effect of this amplitude mismatch on the BER performance will be studied in Section IV-D.
3) When the noise power is small and $A_1 > \sum_{l=2}^{K} A_l|\rho_{1l}|$, i.e., user 1 is sufficiently strong [see the paragraph following (A.7) in Appendix A for the establishment of the latter condition], the second and third terms in (11) are negligible. In this case, the converged AIC weight can be approximated by

$$\hat{A}_{1,o} \approx A_1\Big(1 - 2P_e^{(1)}\Big). \tag{13}$$

Fig. 2 shows numerical examples illustrating the asymptotic mean of the AIC weight (i.e., converged canceler weight) in the

[1]The analyses of a receiver front end using an arbitrary length-N linear filter seems to be too complicated to gain much insight (e.g., for the first stage $m = 1$, it involves integrals of two-dimensional (2-D) Gaussian probability density functions).

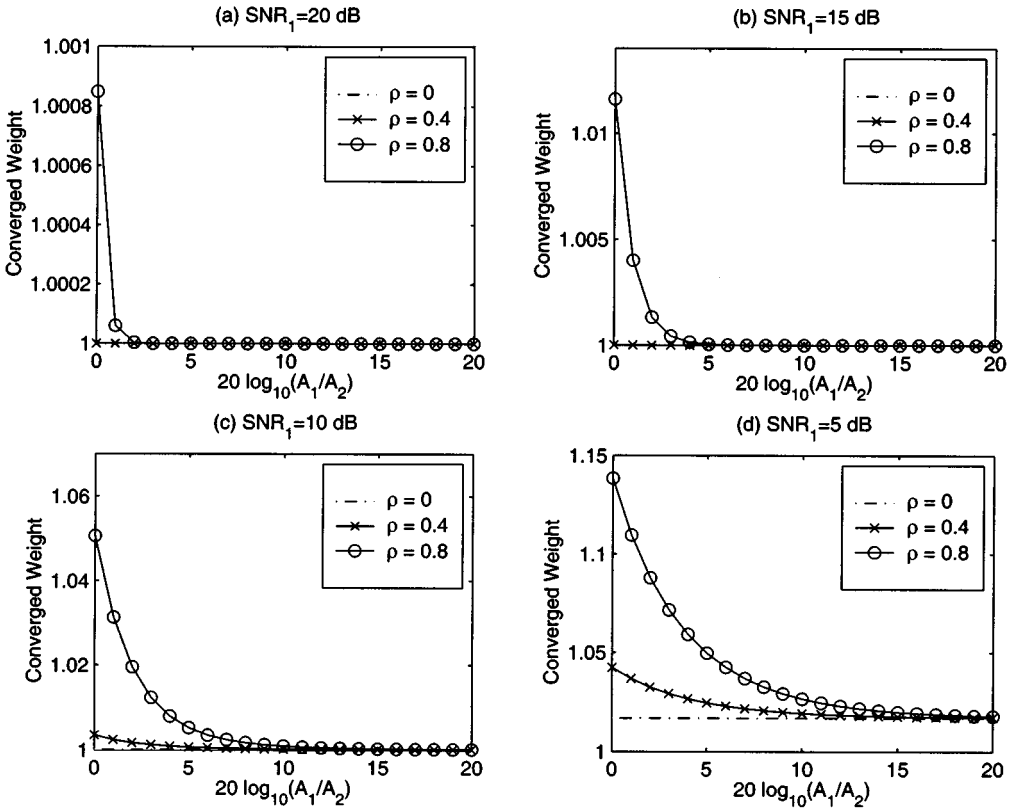

Fig. 2. Converged AIC weight in the first stage ($A_1 = 1$). (a) $\text{SNR}_1 = 20$ dB. (b) $\text{SNR}_1 = 15$ dB. (c) $\text{SNR}_1 = 10$ dB. (d) $\text{SNR}_1 = 5$ dB.

first stage versus the actual received amplitude for the case of $N = 10$ and $K = 2$ such that (11) reduces to

$$\begin{aligned}\hat{A}_{1,o} = A_1\left(1 - 2P_e^{(1)}\right) &+ A_2\rho\left[Q\left(\frac{A_1 - A_2\rho}{\sigma}\right) - Q\left(\frac{A_1 + A_2\rho}{\sigma}\right)\right] \\ &+ \sqrt{\frac{\sigma^2}{2\pi}}\left[\exp\left(\frac{-(A_1 + A_2\rho)^2}{2\sigma^2}\right)\right. \\ &\left. + \exp\left(\frac{-(A_1 - A_2\rho)^2}{2\sigma^2}\right)\right] \qquad (14)\end{aligned}$$

where $\rho \triangleq \rho_{21}$. To illustrate the near–far effect, the received amplitude of the stronger user is normalized to $A_1 = 1$ (recall that we have assumed $A_1 \geq A_2$); the power of the weaker user is varied from 0 to -20 dB with respect to that of user 1, i.e., from $A_2 = A_1$ to $A_2 = 0.1A_1$. The background noise level was set to achieve a specific $\text{SNR}_1 = 20\log_{10}(A_1/\sigma)$ at the MF output. Since (14) is an even function of ρ, $\hat{A}_{1,o}$ is plotted only for $\rho \geq 0$.

Observe that $\hat{A}_{1,o}$ is very close to the actual received amplitude for $\text{SNR}_1 \geq 10$ dB; even for the worst case of $A_2 = 1$ and $\rho = 0.8$, the percentage amplitude mismatch ($|\hat{A}_{1,o} - A_1|/A_1) \times 100\%$ is less than 5%. As the ratio A_1/A_2 increases and ρ becomes smaller, $\hat{A}_{1,o}$ becomes closer to the actual received amplitude. For low SNRs (e.g., $\text{SNR}_1 = 5$ dB), the amplitude mismatch is less than 8%, provided $\rho < 0.6$. Only for extreme conditions with very large MAI (e.g., $A_2 = A_1$ and $\rho = 0.8$) and a high background noise level does the amplitude estimate degrade severely.

B. Asymptotic Mean of the AIC Weight for $m > 1$

Substituting

$$\mathbf{e}_m = \mathbf{r} - \sum_{j=1}^{m-1} \hat{b}_j \hat{A}_{j,o}\mathbf{s}_j \qquad (15)$$

into (10), the asymptotic mean of the AIC weight in the mth stage can be written as

$$\begin{aligned}\hat{A}_{m,o} = &\sum_{j=1}^{K} A_j E\left[\hat{b}_m b_j\right]\rho_{mj} \\ &- \sum_{j=1}^{m-1} \hat{A}_{j,o} E\left[\hat{b}_m \hat{b}_j\right]\rho_{mj} + E\left[\hat{n}_m \hat{b}_m\right]. \qquad (16)\end{aligned}$$

As in the first stage, the AIC weight of the mth ($m > 1$) stage is biased unless the decisions in the current and preceding stages are always correct. Although it is possible to obtain an exact expression for (16), it involves multidimensional integrals of a multivariate Gaussian probability density function (pdf) since $\hat{b}_m$ is, in general, correlated not only with $\{b_j\}_{j=1}^{K}$ and $\hat{n}_m$ but also with previous decisions $\{\hat{b}_j\}_{j=1}^{m-1}$.

In order to gain more insight into the steady-state properties of the ASIC, we make the following "normal-operating" assumptions [25], i.e., the ASIC is operating in an environment

with a substantially high signal-to-interference-plus-noise ratio (SINR) such that the decisions in each stage are sufficiently reliable.

i) $E[\hat{b}_l b_j] \approx 0$ for $l \neq j$.
ii) $E[\hat{b}_l \hat{b}_j] \approx 0$ for $l \neq j$.
iii) $E[\hat{n}_j \hat{b}_l] \approx 0 \; \forall l, j$.

Thus, (16) can be approximated by

$$\begin{aligned} \hat{A}_{m,o} &\approx A_m E\left[\hat{b}_m b_m\right] \\ &= A_m\left(1 - 2P_e^{(m)}\right) \end{aligned} \tag{17}$$

where the last equality follows from the derivation in Appendix A [and was given in (13) for $m = 1$]. The accuracy of this approximation is verified by the computer simulations in Section VI. It was pointed out in [26] that $A_m(1 - 2P_e^{(m)})$ is the (optimal) canceler weight that minimizes the power of the cancellation error $E[||A_m b_m \mathbf{s}_m - \hat{A}_m \hat{b}_m \mathbf{s}_m||^2]$. As a result, the ASIC can be considered as an approximate implementation, in high SINR conditions, of this error criterion.

C. Output Power and Asymptotic Variance of the AIC Weight

The output power of stage m at convergence can be written as

$$\begin{aligned} \xi_o^{(m)} &\triangleq \frac{1}{N} E\left[\mathbf{e}_{m+1}^T \mathbf{e}_{m+1}\right]\Big|_{\hat{A}_m(k) = \hat{A}_{m,o}} \\ &= \xi_o^{(m-1)} - \frac{1}{N}\hat{A}_{m,o}^2 \end{aligned} \tag{18}$$

where we have used $\mathbf{e}_{m+1} = \mathbf{e}_m - \hat{A}_{m,o}\hat{b}_m \mathbf{s}_m$ at convergence and substituted (10). Note that

$$\xi_o^{(0)} = \frac{1}{N} E\left[\mathbf{r}^T \mathbf{r}\right] = \frac{1}{N}\sum_{j=1}^{K} A_j^2 + \sigma^2. \tag{19}$$

Using (13) and (17) for high SINR conditions, we can write

$$\xi_o^{(m)} \approx \frac{1}{N}\left[\sum_{j=1}^{m} 4A_j^2 P_e^{(j)}\left(1 - P_e^{(j)}\right) + \sum_{j=m+1}^{K} A_j^2 + N\sigma^2\right]. \tag{20}$$

Recognizing that $\xi_o^{(m)}$ is equivalent to the minimum MSE achieved by the Wiener solution, the asymptotic variance of the AIC weight in stage m is given by [20]

$$\lim_{k\to\infty} E\left[\left(\hat{A}_m(k) - \hat{A}_{m,o}\right)^2\right] = \frac{\mu \xi_o^{(m)}}{1 - \mu/N}. \tag{21}$$

The asymptotic variance is controlled by the step-size parameter (as expected) and decreases across the stages since $\xi_o^{(m)}$ is decreasing with increasing m according to (18).

D. BER Analysis

Next, we derive the steady-state BER of the ASIC that incorporates amplitude mismatch when the AIC has converged to a biased estimate of the actual amplitude. We employ an approach similar to that described in [5], where the BER of a one-stage PIC was analyzed using perfect knowledge of the received amplitudes. The approach outlined here is easier to formulate because the error probability is obtained by conditioning on the transmitted user data and not on the noise realization as in [5]. Note that the first stage, which is simply a matched filter, does not suffer from the effects of amplitude mismatch; its BER is given by (12).

In order to investigate the steady-state BER of the ASIC, it is assumed that μ is sufficiently small such that the AIC weight at convergence is fixed at the Wiener weight, i.e., $\hat{A}_l(k) = \hat{A}_{l,o}$, yielding a known amplitude mismatch given $\{A_j\}$, $\{\rho_{lj}\}$, and σ. Using (15), the decision variable in the mth stage ($m > 1$) is given by

$$\begin{aligned} z_m &= \mathbf{s}_m^T \mathbf{e}_m \\ &= A_m b_m + \sum_{l=1}^{m-1}\left(A_l b_l - \hat{A}_{l,o}\hat{b}_l\right)\rho_{ml} \\ &\quad + \sum_{l=m+1}^{K} A_l b_l \rho_{ml} + \hat{n}_m \\ &= A_m b_m + \sum_{l=1}^{m-1} A_l\left(b_l - \hat{b}_l\right)\rho_{ml} - \sum_{l=1}^{m-1}\Delta A_l \hat{b}_l \rho_{ml} \\ &\quad + \sum_{l=m+1}^{K} A_l b_l \rho_{ml} + \hat{n}_m \end{aligned} \tag{22}$$

where $\Delta A_l \triangleq \hat{A}_{l,o} - A_l$, and $\hat{b}_l = \text{sgn}[z_l]$. Since the users are detected in order, the BER of user m is affected by the amplitude mismatch in previous stages, i.e., $\Delta\mathbf{A}_m \triangleq [\Delta A_1, \ldots, \Delta A_{m-1}]$. The BER conditioned on $\Delta\mathbf{A}_m$ is

$$\begin{aligned} P_{e|\Delta\mathbf{A}_m}^{(m)} &= P\left(\hat{b}_m \neq b_m | \Delta\mathbf{A}_m\right) \\ &= \sum_{b_m}\sum_{\beta_m}\sum_{\hat{\beta}_m} P\left(\hat{b}_m \neq b_m, b_m, \beta_m, \hat{\beta}_m | \Delta\mathbf{A}_m\right) \end{aligned} \tag{23}$$

where $\beta_m \triangleq [b_1, \ldots, b_{m-1}, b_{m+1}, \ldots, b_K]$ can take on one of 2^{K-1} possible values, and $\hat{\beta}_m \triangleq [\hat{b}_1, \ldots, \hat{b}_{m-1}]$ can take on one of 2^{m-1} possible values. Note that the second and third summations in (23) are performed over all possible values of β_m and $\hat{\beta}_m$, respectively. Using Bayes' rule such that

$$\begin{aligned} &P\left(\hat{b}_m \neq b_m, b_m, \beta_m, \hat{\beta}_m | \Delta\mathbf{A}_m\right) \\ &\quad = P\left(\hat{b}_m \neq b_m, \hat{\beta}_m | b_m, \beta_m, \Delta\mathbf{A}_m\right) P(b_m, \beta_m | \Delta\mathbf{A}_m) \\ &\quad = 2^{-K} P\left(\hat{b}_m \neq b_m, \hat{\beta}_m | b_m, \beta_m, \Delta\mathbf{A}_m\right) \end{aligned} \tag{24}$$

(23) can be rewritten as

$$P_{e|\Delta\mathbf{A}_m}^{(m)} = 2^{-K} \sum_{b_m} \sum_{\beta_m} \sum_{\hat{\beta}_m} \cdot P\Big(\hat{b}_m \neq b_m, \hat{\beta}_m | b_m, \beta_m, \Delta\mathbf{A}_m\Big) \quad (25)$$

where we have used the assumption that the $\{b_l\}$ are i.i.d. random variables taking on the values ± 1 with equal probability. Observe that

$$\begin{aligned} &P\Big(\hat{b}_m \neq b_m, \hat{\beta}_m | b_m, \beta_m, \Delta\mathbf{A}_m\Big) \\ &\quad = P(\hat{n}_m \in R_m, \hat{n}_1 \in R_1, \ldots \\ &\qquad \hat{n}_{m-1} \in R_{m-1} | b_m, \beta_m, \Delta\mathbf{A}_m) \\ &\quad = P(\hat{n}_m \in R_m, \hat{n}_1 \in R_1, \ldots, \hat{n}_{m-1} \in R_{m-1}) \\ &\quad = \int_{R_1} \cdots \int_{R_m} f(\hat{n}_1, \ldots, \hat{n}_m)\, d\hat{n}_1 \cdots d\hat{n}_m \end{aligned} \quad (26)$$

where $f(\hat{n}_1, \ldots, \hat{n}_m)$ is the joint Gaussian pdf of the noise. In order to remove the conditioning in the first equality of (26), we have used the assumption that the $\{\hat{n}_l\}$ are independent of the user data $\{b_l\}$ and $\Delta\mathbf{A}_m$. The regions $\{R_l\}$ are defined as follows. For $l = m$

$$R_m \triangleq \begin{cases} \hat{n}_m < -A_m - \displaystyle\sum_{j=1}^{m-1} A_j\Big(b_j - \hat{b}_j\Big)\rho_{mj} \\ \qquad + \displaystyle\sum_{j=1}^{m-1} \Delta A_j \hat{b}_j \rho_{mj} \\ \qquad - \displaystyle\sum_{j=m+1}^{K} A_j b_j \rho_{mj}, & b_m = 1 \\ \hat{n}_m > A_m - \displaystyle\sum_{j=1}^{m-1} A_j\Big(b_j - \hat{b}_j\Big)\rho_{mj} \\ \qquad + \displaystyle\sum_{j=1}^{m-1} \Delta A_j \hat{b}_j \rho_{mj} \\ \qquad - \displaystyle\sum_{j=m+1}^{K} A_j b_j \rho_{mj}, & b_m = -1 \end{cases} \quad (27)$$

which represents the set of $\{\hat{n}_m\}$ that produces an erroneous decision $\hat{b}_m$ given $b_m, \beta_m, \hat{\beta}_m$, and $\Delta\mathbf{A}_m$. For $l = 1, \ldots, m-1$

$$R_l \triangleq \begin{cases} \hat{n}_l > -A_l b_l - \displaystyle\sum_{j=1}^{l-1} A_j\Big(b_j - \hat{b}_j\Big)\rho_{lj} \\ \qquad + \displaystyle\sum_{j=1}^{l-1} \Delta A_j \hat{b}_j \rho_{lj} - \sum_{j=l+1}^{K} A_j b_j \rho_{lj}, & \hat{b}_l = 1 \\ \hat{n}_l < -A_l b_l - \displaystyle\sum_{j=1}^{l-1} A_j\Big(b_j - \hat{b}_j\Big)\rho_{lj} \\ \qquad + \displaystyle\sum_{j=1}^{l-1} \Delta A_j \hat{b}_j \rho_{lj} - \sum_{j=l+1}^{K} A_j b_j \rho_{lj}, & \hat{b}_l = -1 \end{cases} \quad (28)$$

which denotes the set of $\{\hat{n}_l\}$ that produces a specific decision $\hat{b}_l$ ($+1$ or -1) given $b_l, \beta_l, \hat{\beta}_l$, and $\Delta\mathbf{A}_l$. Inserting (26) into (25) yields

$$P_{e|\Delta\mathbf{A}_m}^{(m)} = 2^{-K} \sum_{b_m} \sum_{\beta_m} \sum_{\hat{\beta}_m} \cdot \int_{R_1} \cdots \int_{R_m} f(\hat{n}_1, \ldots, \hat{n}_m)\, d\hat{n}_1 \cdots d\hat{n}_m \quad (29)$$

which requires 2^{K+m-1} m-dimensional integrations. Note that (29) is the steady-state BER of the ASIC since $\Delta\mathbf{A}_m$ represents the amplitude mismatch of the converged AIC weights. For a random amplitude mismatch $\Delta\mathbf{A}_m$, the overall BER can be obtained by averaging $P_{e|\Delta\mathbf{A}_m}^{(m)}$, i.e.,

$$P_e^{(m)} = \int_{\Delta\mathbf{A}_m} P_{e|\Delta\mathbf{A}_m}^{(m)} f(\Delta\mathbf{A}_m)\, d\Delta\mathbf{A}_m \quad (30)$$

where $f(\Delta\mathbf{A}_m)$ is the joint pdf of the amplitude mismatch.

V. Steady-State Properties of the CSIC

For comparison purposes, we also investigate the corresponding properties of the CSIC based on the matrix approach developed in [18], where the convergence properties of the BER of an iterative linear SIC (i.e., using the magnitude of the MF output as the user amplitude estimate) was studied. The CSIC examined in this paper can be viewed as the first iterative stage of such a structure. As in the previous analyses of the ASIC, we assume that user m is detected in stage m.

First, consider the amplitude estimate of the CSIC. We can write the magnitude of the MF output of stage m as

$$\begin{aligned} \tilde{A}_m &= \left|\mathbf{s}_m^T \mathbf{e}_m\right| \\ &= \tilde{b}_m \mathbf{s}_m^T \tilde{\mathbf{e}}_m \end{aligned} \quad (31)$$

where $\tilde{\mathbf{e}}_m$ and $\tilde{b}_m \triangleq \text{sgn}[\mathbf{s}_m^T \tilde{\mathbf{e}}_m]$ are the input and detected bit, respectively, of the mth stage of the CSIC, and we have used $|x| = x\text{sgn}[x]$ to arrive at the second equality in (31). The tilde notation is used to denote the various quantities/signals of the CSIC (to distinguish them from those of the ASIC).

A. Review of the Matrix Approach

Using this approach, we can express $\tilde{\mathbf{e}}_m$ as a linear matrix filtering operation on the chip MF output $\mathbf{r}$. Setting $\hat{A}_m(k) = \tilde{A}_m$ in Fig. 1 (which reduces to the CSIC in this case), the input signal vector of stage m can be written as

$$\begin{aligned} \tilde{\mathbf{e}}_m &= \tilde{\mathbf{e}}_{m-1} - \tilde{A}_{m-1} \tilde{b}_{m-1} \mathbf{s}_{m-1} \\ &= \tilde{\mathbf{e}}_{m-1} - \mathbf{s}_{m-1} \mathbf{s}_{m-1}^T \tilde{\mathbf{e}}_{m-1} \\ &= \big(\mathbf{I} - \mathbf{s}_{m-1} \mathbf{s}_{m-1}^T\big) \tilde{\mathbf{e}}_{m-1} \end{aligned} \quad (32)$$

where we have substituted $\tilde{A}_{m-1} = \tilde{b}_{m-1} \mathbf{s}_{m-1}^T \tilde{\mathbf{e}}_{m-1}$ [similar to (31)], and $\mathbf{I}$ is the identity matrix of size N. Note that $\mathbf{I} - \mathbf{s}_{m-1}\mathbf{s}_{m-1}^T$ is a projection matrix onto the orthogonal complement of $\mathbf{s}_{m-1}$. Iterating on $\tilde{\mathbf{e}}_{m-1}$ in the same manner gives

$$\tilde{\mathbf{e}}_m = \mathbf{T}_m \mathbf{r} \quad (33)$$

where we have defined the signal transfer matrix $\mathbf{T}_m \triangleq (\mathbf{I} - \mathbf{s}_{m-1}\mathbf{s}_{m-1}^T) \cdots (\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^T)$, which accounts for the composite effect of the previous $m-1$ stages of cancellation. It is clear

that $\mathbf{T}_1 = \mathbf{I}$. Using (4) and (33), the decision variable at stage m can be expressed as

$$\begin{aligned}\tilde{z}_m &= \mathbf{s}_m^T \tilde{\mathbf{e}}_m \\ &= A_m b_m \tilde{\rho}_{mm} + \sum_{j \neq m} A_j b_j \tilde{\rho}_{mj} + \tilde{n}_m \end{aligned} \tag{34}$$

where $\tilde{\rho}_{mj} \triangleq \mathbf{s}_m^T \mathbf{T}_m \mathbf{s}_j$ is the "effective" correlation of the spreading sequences of user m and j after the preceding $m-1$ cancellations, and $\tilde{n}_m \triangleq \mathbf{s}_m^T \mathbf{T}_m \mathbf{n}$ is the noise component at the output of the symbol MF. It is straightforward to show that $\tilde{n}_m$ is a zero-mean Gaussian random variable with variance $\tilde{\sigma}_m^2 \triangleq \sigma^2[\mathbf{s}_m^T \mathbf{T}_m \mathbf{T}_m^T \mathbf{s}_m]$. Note that $\tilde{z}_m$ has the same form as the decision variable of the mth branch of the conventional MF detector, i.e., $A_m b_m + \sum_{j \neq m} A_j b_j \rho_{mj} + \hat{n}_m$. Hence, we may also view the symbol MF operation of stage m in (34) as a despreading of the received signal vector $\mathbf{r}$ with the "modified" spreading sequence $\mathbf{T}_m^T \mathbf{s}_m$. Based on this analogy, the BER of stage m of the CSIC is given by [18]

$$\begin{aligned}\tilde{P}_e^{(m)} = 2^{-(K-1)} &\sum_{b_1} \cdots \sum_{b_{m-1}} \sum_{b_{m+1}} \\ &\cdots \sum_{b_K} Q\left(\frac{A_m \tilde{\rho}_{mm} + \sum_{l \neq m} A_l b_l \tilde{\rho}_{ml}}{\tilde{\sigma}_m} \right) \end{aligned} \tag{35}$$

which has a form similar to that of the MF detector in (12).

B. Properties of $\mathbf{T}_m$ and $\tilde{\rho}_{mj}$

It will be useful to examine the properties of $\mathbf{T}_m$ and $\tilde{\rho}_{mj}$ for $m > 1$ ($m = 1$ is trivial since $\tilde{\rho}_{1j} = \rho_{1j}\ \forall\, j$). First, note that $\mathbf{T}_m$ is symmetric and idempotent if and only if the $\{\mathbf{s}_j\}_{j=1}^{m-1}$ are mutually orthogonal. (The trivial case where $\{\mathbf{s}_j\}_{j=1}^{m-1}$ are collinear is not considered here because the users' spreading sequences cannot be collinear in practice.) Therefore, $\mathbf{T}_m$ is generally not an orthogonal projection matrix, even though it is a product of orthogonal projection matrices $\{\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^T\}$. In addition, note that $\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^T$ is positive semi-definite (PSD), which can easily be shown by expanding $\mathbf{x}^T(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^T)\mathbf{x}$ and applying the Cauchy–Schwarz inequality. Consequently, $\mathbf{T}_m$ is PSD since it is a product of PSD matrices [27]. Second, $\mathbf{T}_m \mathbf{s}_1 = \mathbf{0}$ because $(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^T)\mathbf{s}_1 = \mathbf{0}$. As a result, $\tilde{\rho}_{m1} = 0$ so that stage m ($m > 1$) is free of the MAI due to user 1 (i.e., user 1 is completely removed in the first stage). However, the other users are not completely removed (i.e., residual cancellation error exists) since $\tilde{\rho}_{mj}$ is generally nonzero for $j \neq 1$. Third, $\tilde{\rho}_{mm} = 1$ if and only if the $\{\mathbf{s}_j\}_{j=1}^{m-1}$ are mutually orthogonal; otherwise, $0 \leq \tilde{\rho}_{mm} < 1$. The following is a sketch of the proof. The non-negativity of $\tilde{\rho}_{mm} = \mathbf{s}_m^T \mathbf{T}_m \mathbf{s}_m$ follows immediately because $\mathbf{T}_m$ is PSD. To see that $\tilde{\rho}_{mm}$ is upper-bounded by unity, observe that $\|\mathbf{T}_m \mathbf{s}_m\| \leq \|\mathbf{s}_m\| = 1$, i.e., the length of the vector after projection is no greater than that before projection. Equality is achieved if and only if the $\{\mathbf{s}_j\}_{j=1}^{m-1}$ are mutually orthogonal. From the Cauchy–Schwarz inequality, we have that

$$\tilde{\rho}_{mm} = \mathbf{s}_m^T \mathbf{T}_m \mathbf{s}_m \leq \|\mathbf{s}_m\| \|\mathbf{T}_m \mathbf{s}_m\| = \|\mathbf{T}_m \mathbf{s}_m\| \leq 1. \tag{36}$$

The second equality holds if and only if $\mathbf{T}_m \mathbf{s}_m$ and $\mathbf{s}_m$ are collinear (or, equivalently, the $\{\mathbf{s}_j\}_{j=1}^{m-1}$ are mutually orthogonal, which is the same condition needed to achieve the last equality). Note that the partial signal cancellation effects (i.e., portions of user m's signal are canceled prior to stage m) result in $\tilde{\rho}_{mm} < 1$. We will demonstrate that the accumulation of partial signal cancellation effects and residual cancellation errors causes relatively large biases in the amplitude estimates of the CSIC.

C. Mean, Variance, and Output Power

From (31) and (34), the mean of the amplitude estimate $\tilde{A}_m$ can be written as

$$\begin{aligned} E\left[\tilde{A}_m\right] &= E\left[\tilde{b}_m \mathbf{s}_m^T \tilde{\mathbf{e}}_m\right] \\ &= A_m \tilde{\rho}_{mm} E\left[b_m \tilde{b}_m\right] \\ &\quad + \sum_{j \neq m} A_j \tilde{\rho}_{mj} E\left[b_j \tilde{b}_m\right] + E\left[\tilde{n}_m \tilde{b}_m\right]. \end{aligned} \tag{37}$$

Recognizing that (37) has the same form as the first equality in (11), a closed-form expression for $E[\tilde{A}_m]$ can be obtained by directly applying the results in Appendix A, yielding

$$\begin{aligned} E\left[\tilde{A}_m\right] = A_m \tilde{\rho}_{mm} &\left(1 - 2\tilde{P}_e^{(m)}\right) + 2^{-(K-1)} \\ &\cdot \sum_{j \neq m} A_j \tilde{\rho}_{mj} \sum_{b_1} \cdots \sum_{b_{j-1}} \sum_{b_{j+1}} \cdots \sum_{b_K} \\ &\cdot \left[Q\left(\frac{-A_j \tilde{\rho}_{mj} + \sum_{l \neq j} A_l b_l \tilde{\rho}_{ml}}{\tilde{\sigma}_m} \right) \right. \\ &\quad \left. - Q\left(\frac{A_j \tilde{\rho}_{mj} + \sum_{l \neq j} A_l b_l \tilde{\rho}_{ml}}{\tilde{\sigma}_m} \right) \right] \\ &+ 2^{-K} \sqrt{\frac{2\tilde{\sigma}_m^2}{\pi}} \sum_{b_1} \cdots \sum_{b_K} \\ &\cdot \exp\left(\frac{-\left(\sum_{l=1}^{K} A_l b_l \tilde{\rho}_{ml} \right)^2}{2\tilde{\sigma}_m^2} \right) \end{aligned} \tag{38}$$

where $\tilde{P}_e^{(m)}$ is the error probability of the mth stage of the CSIC given in (35).

The second moment and variance of $\tilde{A}_m$, which are derived in Appendix B, are given by

$$E\left[\tilde{A}_m^2\right] = \sum_{l=1}^{K} A_l^2 \tilde{\rho}_{ml}^2 + \tilde{\sigma}_m^2 \tag{39}$$

and

$$\text{var}\left[\tilde{A}_m\right] = E\left[\tilde{A}_m^2\right] - E^2\left[\tilde{A}_m\right] \tag{40}$$

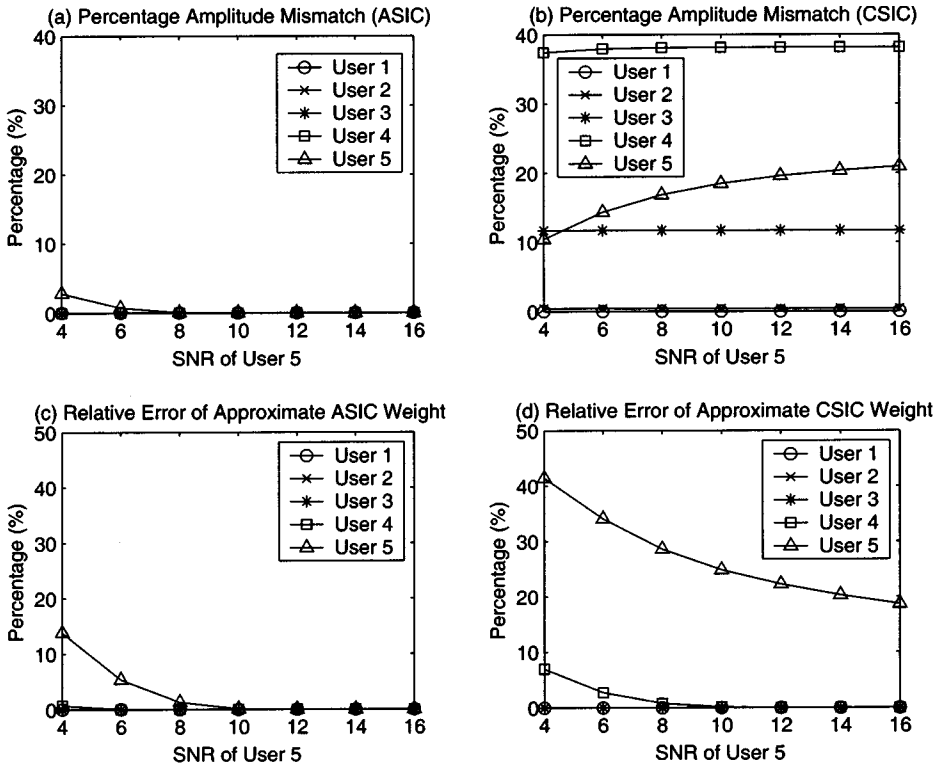

Fig. 3. Amplitude estimates of the ASIC and CSIC for $K = 5$ and power profile $[20, 15, 10, 5, 0]$ dB. (a) Percentage amplitude mismatch of the ASIC. (b) Percentage amplitude mismatch of the CSIC. (c) Relative error of approximating the converged AIC weight $\hat{A}_{m,o}$ by $A_m(1-2P_e^{(m)})$. (d) Relative error of approximating $E[\tilde{A}_m]$ by $A_m\tilde{\rho}_{mm}(1-2\tilde{P}_e^{(m)})$.

respectively. Using $\tilde{\mathbf{e}}_{m+1} = \tilde{\mathbf{e}}_m - \tilde{A}_m\tilde{b}_m\mathbf{s}_m$ and (31), the output power of the mth stage of the CSIC can be written as

$$\begin{aligned}\tilde{\xi}^{(m)} &= \frac{1}{N}E\left[\tilde{\mathbf{e}}_{m+1}^T\tilde{\mathbf{e}}_{m+1}\right] \\ &= \tilde{\xi}^{(m-1)} - \frac{1}{N}E\left[\tilde{A}_m^2\right] \end{aligned} \tag{41}$$

with $\tilde{\xi}^{(0)} = (1/N)E[\mathbf{r}^T\mathbf{r}] = \xi_o^{(0)}$.

Using arguments similar to those leading to (13), we see that if the noise power is small and $A_m\tilde{\rho}_{mm} > \sum_{l\neq m} A_l|\tilde{\rho}_{ml}|$ (this inequality corresponds to the open-eye condition), then the second and third terms in (38) are negligible. In this case, we have

$$E\left[\tilde{A}_m\right] \approx A_m\tilde{\rho}_{mm}\left(1-2\tilde{P}_e^{(m)}\right). \tag{42}$$

Thus, (40) can be approximated by

$$\operatorname{var}\left[\tilde{A}_m\right] \approx 4A_m^2\tilde{\rho}_{mm}^2\tilde{P}_e^{(m)}\left(1-\tilde{P}_e^{(m)}\right) + \sum_{l\neq m} A_l^2\tilde{\rho}_{ml}^2 + \tilde{\sigma}_m^2. \tag{43}$$

From these results, we draw the following conclusions.

1) It can easily be seen from (37) and (38) that $E[\tilde{A}_m] = A_m\tilde{\rho}_{mm}$ when no errors occur [i.e., $\tilde{P}_e^{(m)} = 0$ or, equivalently, $\tilde{\sigma}_m = 0$ and $A_m\tilde{\rho}_{mm} > \sum_{l\neq m} A_l|\tilde{\rho}_{ml}|$] since $\tilde{b}_m = b_m$ is uncorrelated with $\{b_j\}_{j\neq m}$ and $\tilde{n}_m$. Even in this ideal case, unless orthogonal spreading sequences are employed, the amplitude estimates of the CSIC for $m > 1$ are still biased because $\tilde{\rho}_{mm}$ is less than unity. As will be illustrated using numerical examples in the next section, this bias can be relatively large in near–far situations and could cause a serious loss in performance for the CSIC. In contrast, the asymptotic mean of an ASIC stage is unbiased if the decisions up to and including that stage are always correct.
2) For $m = 1$, the mean of the amplitude estimate is identical to the asymptotic mean of the ASIC amplitude estimate, i.e., $E[\tilde{A}_1] = \lim_{k\to\infty} E[\hat{A}_1(k)] = \hat{A}_{1,o}$. However, these two amplitude estimates differ for $m > 1$. This can be seen from the asymptotic results: $\hat{A}_{m,o} \to A_m(1-2P_e^{(m)})$, whereas $E[\tilde{A}_m] \to A_m\tilde{\rho}_{mm}(1-2\tilde{P}_e^{(m)})$ if the open-eye condition is met. The variances of the two amplitude estimates also differ quite significantly, as will be shown via the computer simulations in the next section.

VI. Computer Simulations

A. Amplitude Estimation

1) Mean of the Converged Canceler Weight: In the following example, we compare the bias of the converged AIC weight with that of the amplitude estimate computed for the CSIC. Fig. 3(a) and (b) shows the percentage amplitude mismatch of the ASIC and CSIC, respectively, for a $K = 5$ user system. This percentage is defined as $|Y - A_m|/A_m \times 100\%$, where $Y = \hat{A}_{m,o}$ for the ASIC, and $Y = E[\tilde{A}_m]$ for the CSIC. Note that $\hat{A}_{m,o}$ is approximated by an ensemble average of the converged AIC weight for 10^5 symbols using Monte Carlo simulations with $\mu = 5\times10^{-4}$, whereas $E[\tilde{A}_m]$ is obtained by numerically evaluating (38). The system has four strong users and one weak user: users 1, 2, 3, and 4 are 20, 15, 10, and 5 dB stronger than user 5, respectively (i.e., power profile $= [20, 15, 10, 5, 0]$

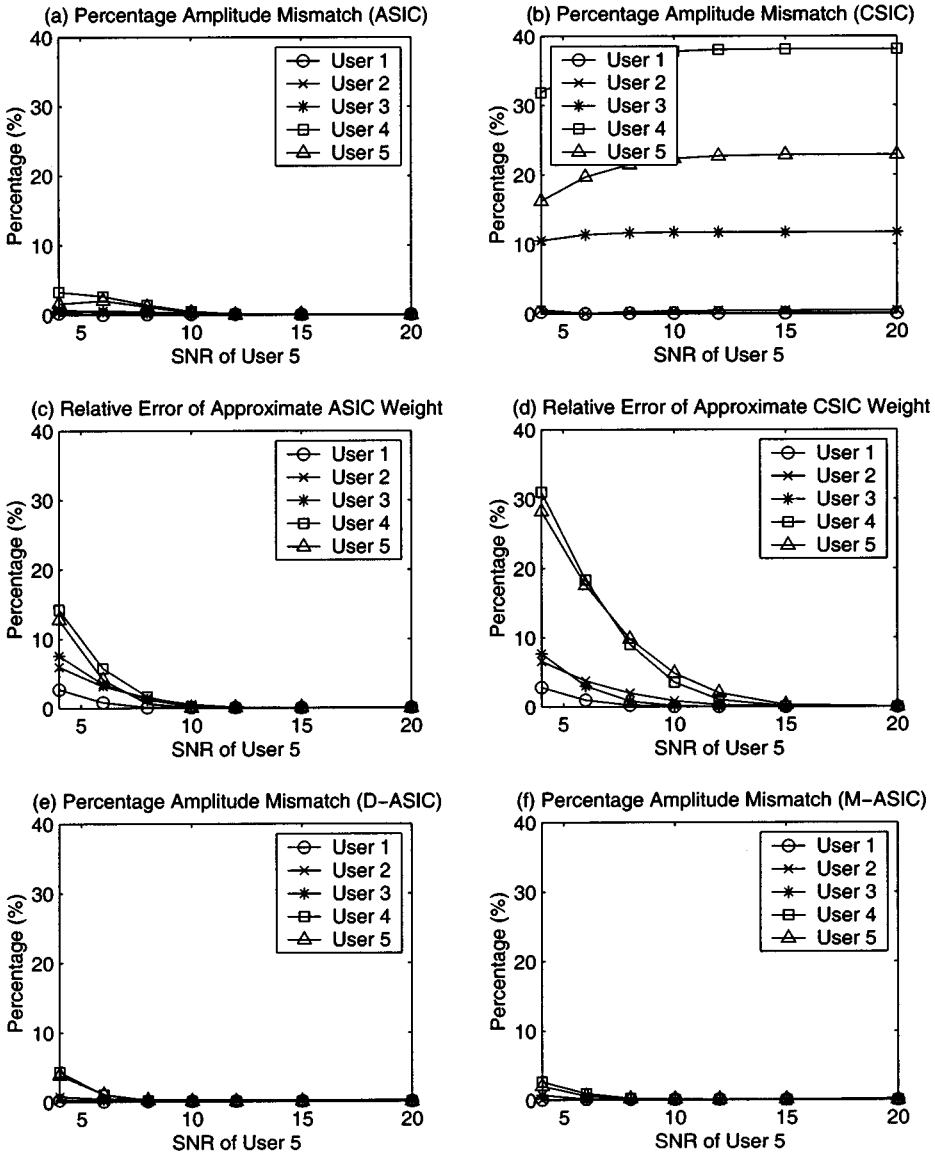

Fig. 4. Amplitude estimates of the ASIC, D-ASIC, M-ASIC, and CSIC for $K = 5$ and power profile $[5,\ 5,\ 3,\ 0,\ 0]$ dB. (a) Percentage amplitude mismatch of the ASIC. (b) Percentage amplitude mismatch of the CSIC. (c) Relative error of approximating the converged AIC weight $\hat{A}_{m,o}$ by $A_m(1-2P_e^{(m)})$. (d) Relative error of approximating $E[\bar{A}_m]$ by $A_m\bar{\rho}_{mm}(1-2\bar{P}_e^{(m)})$. (e) Percentage amplitude mismatch of the D-ASIC. (f) Percentage amplitude mismatch of the M-ASIC.

dB). The amplitude of user 5 is fixed at $A_5 = 1$; the noise power is varied to achieve a specific $\text{SNR}_5 \triangleq 20\log_{10}(A_5/\sigma)$ (this represents the SNR per bit, i.e., the SNR after despreading). The processing gain of the system is $N = 15$, and the correlations between the spreading sequences are

$$\mathbf{R} = \frac{1}{15}\begin{bmatrix} 15 & -1 & -1 & 5 & 5 \\ -1 & 15 & -5 & 5 & 5 \\ -1 & -5 & 15 & -7 & -3 \\ 5 & 5 & -7 & 15 & 3 \\ 5 & 5 & -3 & 3 & 15 \end{bmatrix} \tag{44}$$

where $[\mathbf{R}]_{jl} = \rho_{jl}$.

Observe that the amplitude mismatch of the CSIC is quite large for users 3, 4, and 5. In this example, $\tilde{\rho}_{22} = 0.9956$, $\tilde{\rho}_{33} = 0.8830$, $\tilde{\rho}_{44} = 0.6183$, and $\tilde{\rho}_{55} = 0.7700$ [recall from (38) that the amount of mismatch is dependent on $\tilde{\rho}_{mm}$]. The biases in stages 3, 4, and 5 do not vanish as the noise power decreases. On the other hand, the amplitude mismatch of the ASIC is very small for moderate-to-high SNR_5 and quickly approaches zero as SNR_5 increases. Fig. 3(c) demonstrates the accuracy when approximating $\hat{A}_{m,o}$ by $A_m(1-2P_e^{(m)})$. It indicates that for $\text{SNR}_5 \geq 8$ dB, the converged AIC weights closely approximate the (optimal) canceler weights that minimize the power of the cancellation error. Similarly, Fig. 3(d) illustrates that (42) is an accurate approximation of the mean of the CSIC canceler weight for users 1–4 for $\text{SNR}_5 \geq 8$ dB. Note, in this example, that user 5 does not satisfy the open-eye condition $A_m\tilde{\rho}_{mm} > \sum_{l\neq m} A_l|\tilde{\rho}_{ml}|$; thus, (42) is not valid. This could happen to weaker users in the presence of several very strong users, as in this example.

Fig. 4 shows a second example, illustrating a moderate near–far condition in which the scenario is the same as that of Fig. 3, except that users 1, 2, 3, and 4 are 5, 5, 3, and 0 dB, respectively, stronger than user 5 (i.e., power profile

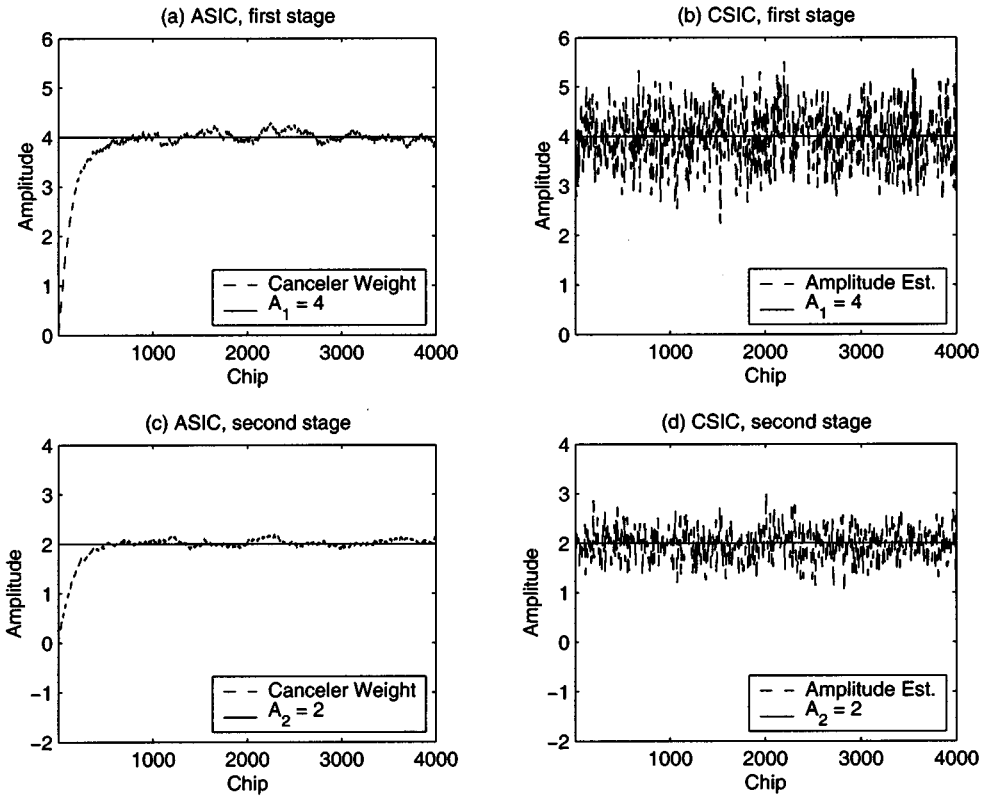

Fig. 5. Amplitude estimates for an AWGN channel. (a) First stage of ASIC (user 1). (b) First stage of CSIC (user 1). (c) Second stage of ASIC (user 2). (d) Second stage of CSIC (user 2).

$= [5, 5, 3, 0, 0]$ dB). Again, it can be seen that the amplitude mismatch of the ASIC is smaller than that of the CSIC, especially for weaker users. Fig. 4(c) and (d) shows the regions where (17) and (42) accurately predict the mean of the amplitude estimates for the ASIC and CSIC, respectively. Fig. 4(e) and (f) plots the percentage amplitude mismatch when the conventional MF detectors in the ASIC are replaced by decorrelators and MMSE receivers, which are referred to as the D-ASIC and M-ASIC, respectively. In stage m, the decorrelator is designed to completely null users $m+1, \ldots, K$;[2] the MMSE receiver in each stage is trained using 1000 symbols. As expected, the latter two detectors yield improved decision-driven amplitude estimates because the symbol estimates are more accurate but at the cost of an increased complexity.

2) Variance, Convergence, and Tracking of the Canceler Weight: Next, we compare the variances of the amplitude estimates. Consider a near–far scenario for a DS/CDMA system with $K = 3$ users and a processing gain of $N = 7$. The received amplitudes are $A_1 = 4$, $A_2 = 2$, and $A_3 = 1$, the spreading sequence correlations are $\rho_{12} = 1/7$, $\rho_{13} = 3/7$, and $\rho_{23} = 1/7$, and the noise power is $\sigma^2 = 0.1$. The AIC weight is initialized to zero, and $\mu = 0.025$ for the first two stages. The weight trajectory in the first stage for a single run of the ASIC is shown in Fig. 5(a), along with the actual received amplitude (which is $A_1 = 4$). The simulation results show that the converged weight (after sample 600) is very close to the received amplitude of user 1.

For comparison purposes, the received amplitude estimate in the first stage of the CSIC for the same signal conditions is shown in Fig. 5(b). Since the received amplitude estimate of the CSIC is obtained from the magnitude of the MF output, it is held constant for $N = 7$ consecutive chips. This suggests that the amplitude estimate should have a smaller variance in the ASIC than in the CSIC, as verified in the plots. Fig. 5(c) and (d) show the weight trajectories for the second stage of the ASIC and the CSIC, respectively. Again, the received amplitude estimate of the ASIC has a smaller variance than that of the CSIC.

Finally, we illustrate the convergence and tracking performance of the AIC for a single-path Rayleigh fading channel with a normalized Doppler[3] of 1.7×10^{-3}. The scenario is the same as that in Fig. 5, except that $A_1 = 2$, $A_2 = \sqrt{2}$, $A_3 = 1$, and $\mu = 0.06$. In this scenario, the received signal in (1) is modified as follows:

$$r(t) = \sum_{j=1}^{K} \sum_{i=-\infty}^{\infty} \gamma_j(t) A_j b_j(i) s_j(t - iT) + n(t) \qquad (45)$$

where $\gamma_j(t)$ is the Rayleigh fading process of user j. The fading coefficients $\{\gamma_j(k) = \gamma_j(t)|_{t=kT_c}\}$ are generated using the model in [28]. For convenience, we set $E[\gamma_j^2(k)] = 1$ and refer to $A_j\gamma_j(k)$ as the received amplitude of user j. Fig. 6(a) and

[2]From the zero-forcing property of the decorrelator, it can be seen that if the decorrelator front end of stage m is designed to null users $1, \ldots, m-1, m+1, \ldots, K$, then the resulting structure has essentially the same performance as that without interference cancellation in [21]. In other words, interference cancellation is redundant in such a design.

[3]The normalized Doppler, which characterizes the fading rate, is defined as the product of the Doppler shift $(v/c)f_c$ and the symbol (bit) interval T, where v is the velocity of the mobile, c is the velocity of light, and f_c is the carrier frequency.

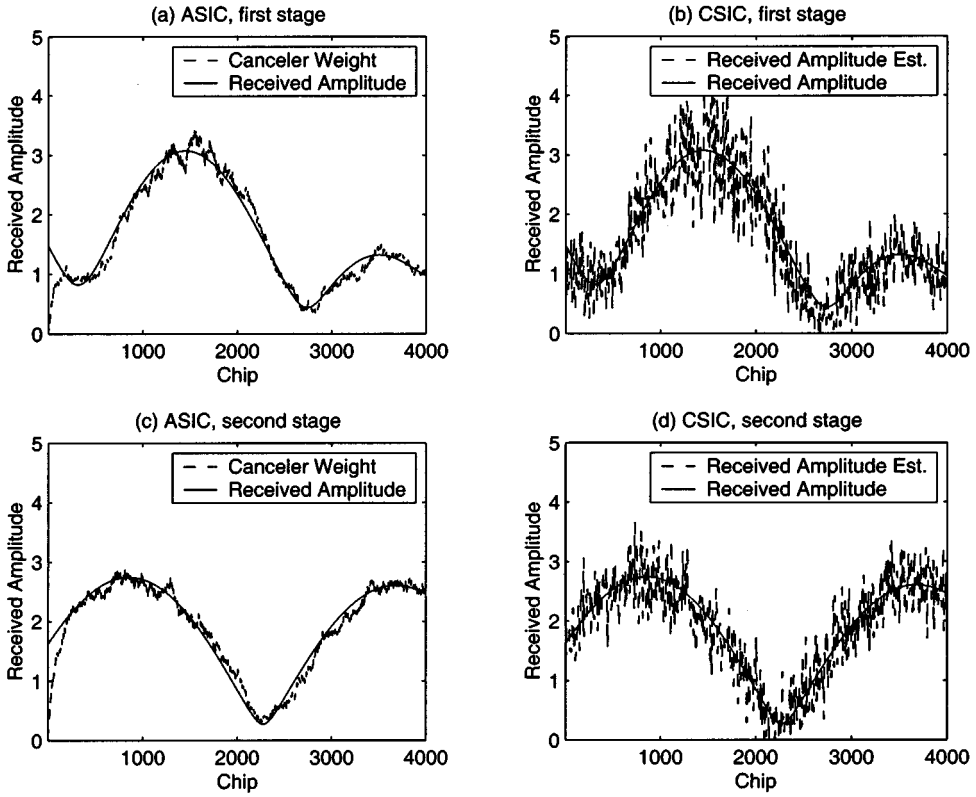

Fig. 6. Received amplitude estimates for a single-path Rayleigh fading channel with a normalized Doppler of 1.7×10^{-3}. (a) First stage of ASIC (user 1). (b) First stage of CSIC (user 1). (c) Second stage of ASIC (user 2). (d) Second stage of CSIC (user 2).

(c) show the weight trajectories in the first and second stages, respectively, for a single run of the ASIC along with the corresponding received amplitudes. Observe that the convergence of the AIC weight is quite fast and that it tracks the received amplitude better than the CSIC in Fig. 6(b) and (d).

B. BER Performance

1) Verification of the BER Analysis: We first verify the accuracy of the steady-state BER analysis using Monte Carlo simulations. Consider a two-user DS/CDMA system with a processing gain of $N = 10$. The amplitude of the stronger user (user 1) is fixed at $A_1 = 1$ with a signal-to-noise ratio of $\mathrm{SNR}_1 = 20\log_{10} A_1/\sigma$, whereas the power of user 2 is varied from 0 to -20 dB with respect to user 1. Fig. 7 shows the ASIC BER curves, along with the corresponding simulation results for $\mu = 0.005$, obtained by numerically evaluating (29) (via Mathematica) with the following amplitude mismatch [see (14)]:

$$\begin{aligned}\Delta A_1 =& -2A_1P_e^{(1)} + A_2\rho \\ &\cdot\left[Q\left(\frac{A_1 - A_2\rho}{\sigma}\right) - Q\left(\frac{A_1 + A_2\rho}{\sigma}\right)\right] \\ &+ \sqrt{\frac{\sigma^2}{2\pi}}\left[\exp\left(\frac{-(A_1 + A_2\rho)^2}{2\sigma^2}\right)\right. \\ &\left.+ \exp\left(\frac{-(A_1 - A_2\rho)^2}{2\sigma^2}\right)\right] \qquad (46)\end{aligned}$$

for $\rho = 0.4$ and four values of SNR_1. It is clear that the analysis closely matches the simulation results.

2) BER Versus Power Ratio: Fig. 8(a) compares the BER versus interferer power of the ASIC with that of the CSIC, the decorrelator, and the conventional MF detector. The system consists of four equally strong users and one weak user: $A_1 = A_2 = A_3 = A_4 \geq A_5$. The power ratio (PR), which is defined by $20\log_{10} A_i/A_5$ for $i = 1, \ldots, 4$, varies from 0 to 20 dB; A_5 is fixed at unity with $\mathrm{SNR}_5 \triangleq 20\log_{10}(A_5/\sigma) = 10$ dB. The processing gain of the system is $N = 15$, and the correlations between the spreading sequences are specified in (44). In each symbol interval, the user to be canceled in a stage of the ASIC and CSIC is determined by selecting the one with the largest magnitude at the MF output in that stage. Since we are

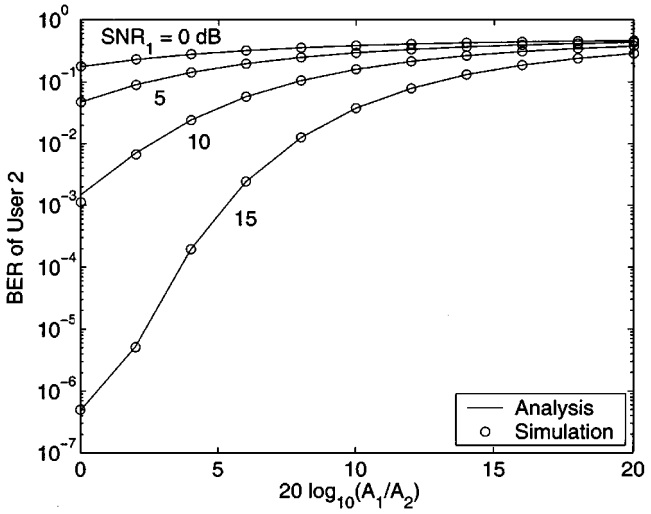

Fig. 7. Steady-state BER for the ASIC. Analysis versus computer simulations for $K = 2$, $N = 10$, $\rho = 0.4$, and $\mu = 0.005$.

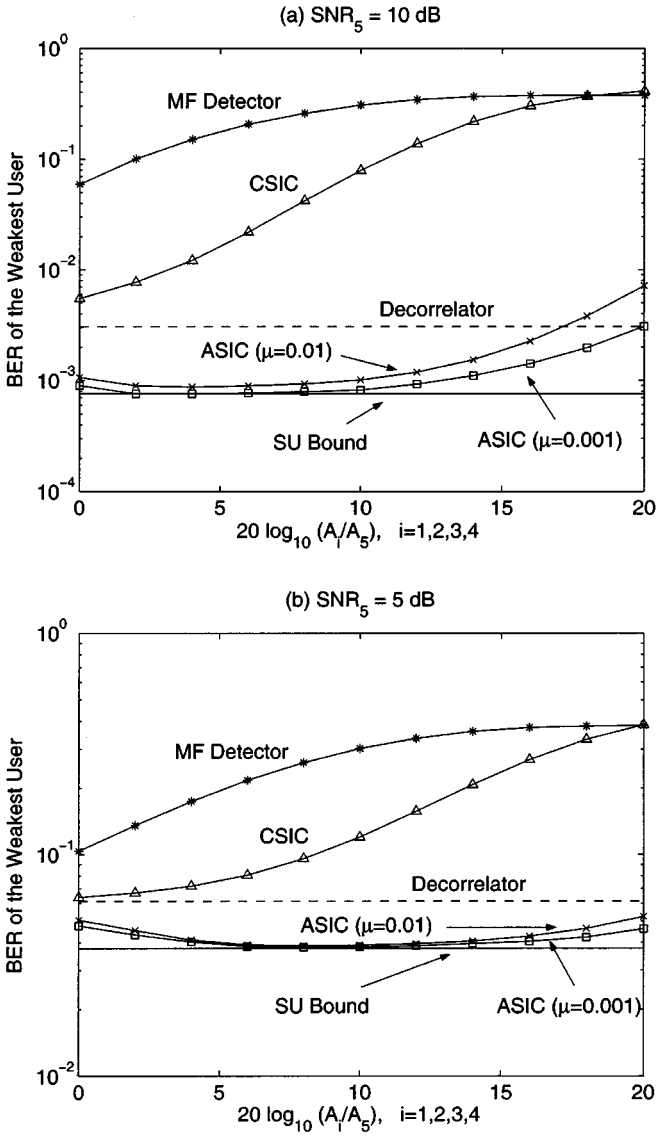

Fig. 8. BER versus power ratio of the ASIC, CSIC, decorrelator, and MF detector. (a) $\mathrm{SNR}_5 = 10$ dB. (b) $\mathrm{SNR}_5 = 5$ dB.

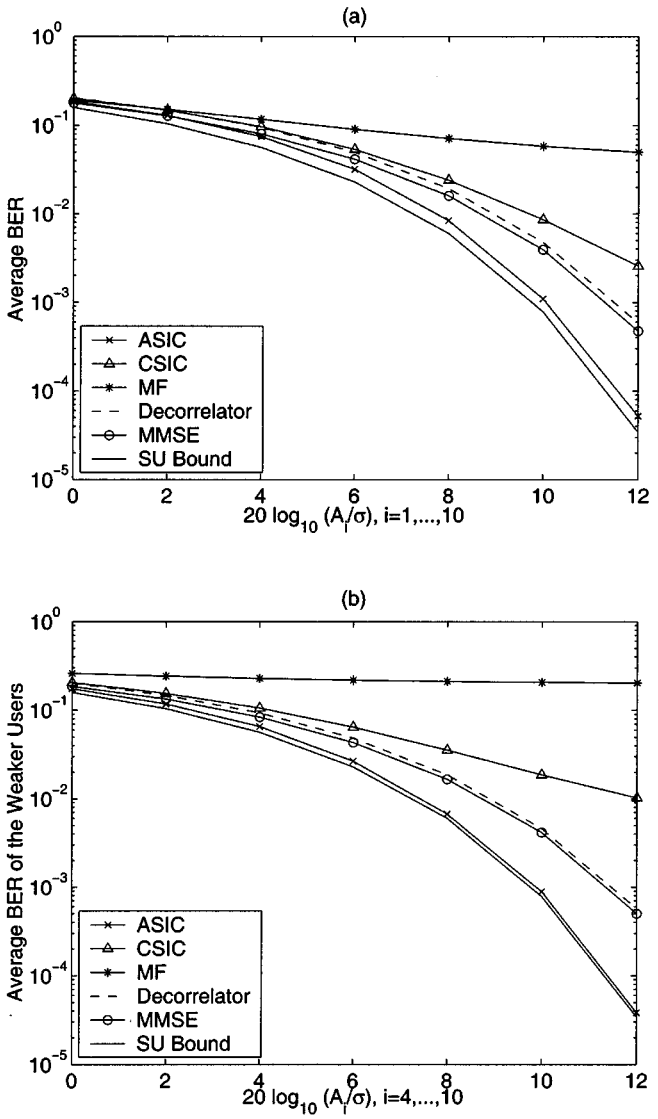

Fig. 9. BER versus SNR of the ASIC, CSIC, decorrelator, MMSE receiver, and MF detector for an AWGN channel. (a) Equal power users. (b) Near–far situation.

primarily interested in near–far scenarios, only the BER curves of the weakest user are plotted.

It can be seen that the CSIC considerably outperforms the MF detector for PR $\leq$ 10 dB. However, the performance gain becomes marginal as the PR increases because the variances of the amplitude estimates increase accordingly [see (39) and (40)]. For large PRs, even a moderate amount of amplitude mismatch from stronger users in the preceding stages can accumulate and cause a considerable amount of residual MAI for the weak user. For the ASIC, the variances of the amplitude estimates also increase with increasing PR [via the increased output power $\xi_o^{(m)}$ in each stage; see (18), (19), and (21)]. However, this increase is greatly deemphasized via the multiplication factor $\mu/(1-\mu/N)$ in (21), where μ is usually small in practice, resulting in a much lower BER compared with that of the CSIC. For PR < 2 dB, the ASIC BER for user 5 improves as the PR increases because the symbol decisions on users 1–4 become more reliable as the SINR of each stage increases, leading to better cancellations. For $2 \leq$ PR < 8 dB, the decisions on users 1–4 are very reliable (no errors occurred in 10^7 symbols in our computer simulations); this result, along with the accurate amplitude estimates provided by the AIC, ensure that the MAI due to users 1–4 is almost completely removed. Therefore, the BER of the weakest user nearly achieves the single-user (SU) bound for BPSK signaling [1].

For PR $>$ 10 dB, the increasing variances of the amplitude estimates become detrimental to the weaker user, resulting in a noticeable increase in the BER with respect to the SU bound. A smaller step-size parameter can be employed to mitigate this problem but at the cost of slower convergence. Observe that the ASIC BER performance for the weaker user is superior to that of the decorrelator for PR $\leq$ 16 dB but falls short as the PR increases due to the increased variance of the amplitude estimates. Fig. 8(b) shows the BER curves for the same scenario as that of Fig. 8(a), except that $\mathrm{SNR}_5 = 5$ dB. Note that in this case, the noise enhancement effect on the decorrelator is more pronounced in low SNR situations; the ASIC still performs better than the decorrelator for PR $= 20$ dB.

3) BER Versus SNR: Fig. 9(a) and (b) shows the BER versus SNR of the ASIC for a $K = 10$ user system and an AWGN

channel under ideal power control (i.e., all users have the same power) and a near–far condition, respectively. The step-size parameter is $\mu = 5\times 10^{-4}$, the processing gain is $N = 31$, and the correlation matrix of the spreading sequences is given in (47), shown at the bottom of the page. In the near–far scenario, there are three strong users, each of which is 10 dB stronger than the remaining users. It can be seen that the performance of the ASIC is very close to the SU bound, and it outperforms the other receivers considered. In addition, observe that the performance of the CSIC receiver degrades significantly for the near–far case because of the poor amplitude estimates.

Fig. 10(a) and (b) show the BER curves of the ASIC for the same scenario as in Fig. 9 for a single-path Rayleigh fading channel where each user undergoes a normalized Doppler of 1.2×10^{-3}. The step-size parameter is $\mu = 0.025$ and $\mu = 0.1$, respectively, for Fig. 10(a) and (b). Note that although the ASIC still significantly outperforms the CSIC and MF receivers, the performance gap between the ASIC and the SU bound is much larger for the fading channel than for the AWGN channel because error-propagation effects are more severe in the former case—especially during deep fades.

4) BER Versus System Load: Fig. 11(a) and (b) illustrate the BER versus the system load of the ASIC, i.e., the number of active users, for an AWGN channel under ideal power control and for a near–far condition, respectively. In the near–far scenario, there are two groups of users such that those in the same group have equal power. The number of users in the first group is fixed at two, whereas that in the second group is varied to change the system load. Users in the first group are 10 dB stronger than those in the second group; each user in the second group has an SNR per bit of 10 dB. The processing gain is $N = 15$, and the step-size parameter is $\mu = 5 \times 10^{-4}$. A user's spreading sequence is randomly generated, and it varies from symbol to symbol. It can be seen that for a low system load, the performance of the ASIC is close to the SU bound; however, it deteriorates rapidly when the system load is increased because of error-propagation effects (the MF front end performs poorly in this case). In addition, note that the performance of the CSIC is similar to that of the conventional MF detector for a high system load.

5) Comparison of Different Receiver Front Ends: Finally, Fig. 12 shows the BER of user 5 versus SNR of the ASIC structure with different front ends for the same scenario used in Fig. 4. As expected, the MMSE receiver and decorrelator front ends provide enhanced BER performance over when the conventional MF detector is used. In addition, the ASIC outperforms the SIC structure in [10] for this example. Note that the performance gap between M-ASIC and the SIC in [10] (both use MMSE receiver front ends but with different amplitude estimators) indicates that the AIC provides more accurate amplitude estimates than does the linear canceler.

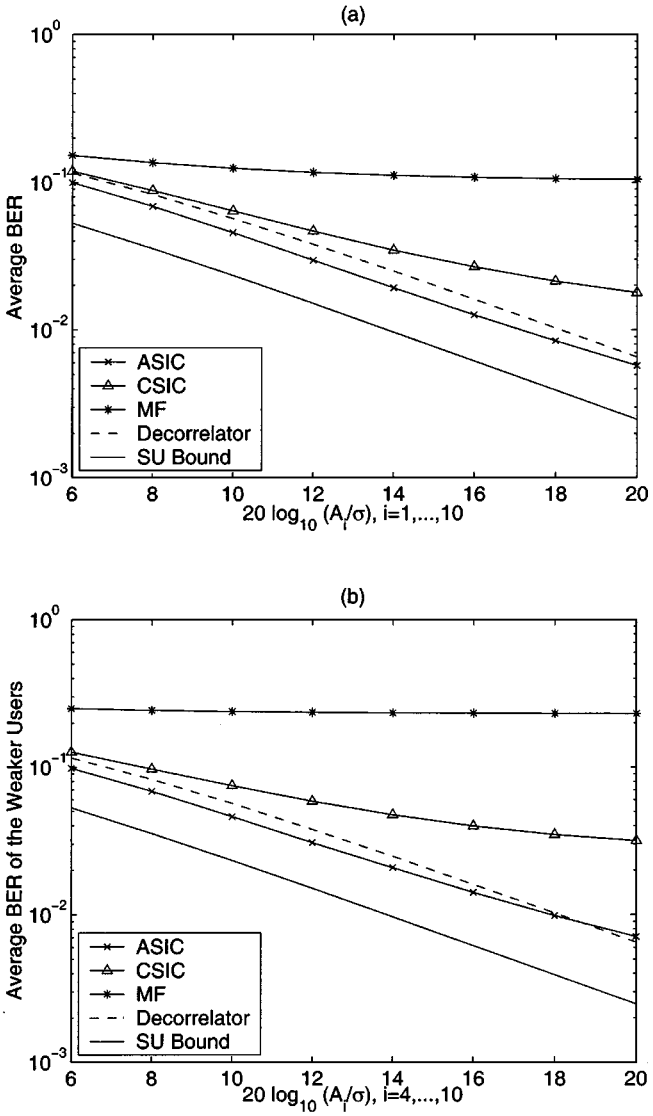

Fig. 10. BER versus SNR of the ASIC, CSIC, decorrelator, and MF detector for a single-path Rayleigh fading channel. (a) Users have equal transmitted power. (b) Near–far situation.

$$\mathbf{R} = \frac{1}{31}\begin{bmatrix} 31 & 7 & 3 & 1 & -1 & -1 & 7 & -5 & 13 & 1 \\ 7 & 31 & 3 & 1 & -1 & 7 & 3 & 7 & 13 & 5 \\ 3 & 3 & 31 & -15 & 3 & -1 & -9 & -1 & 5 & -3 \\ 1 & 1 & -15 & 31 & -3 & -3 & 1 & 5 & -1 & 11 \\ -1 & -1 & 3 & -3 & 31 & 3 & -1 & -1 & 9 & -3 \\ -1 & 7 & -1 & -3 & 3 & 31 & -1 & -5 & -3 & 9 \\ 7 & 3 & -9 & 1 & -1 & -1 & 31 & -1 & 5 & -7 \\ -5 & 7 & -1 & 5 & -1 & -5 & -1 & 31 & 1 & 5 \\ 13 & 13 & 5 & -1 & 9 & -3 & 5 & 1 & 31 & -1 \\ 1 & 5 & -3 & 11 & -3 & 9 & -7 & 5 & -1 & 31 \end{bmatrix} \tag{47}$$

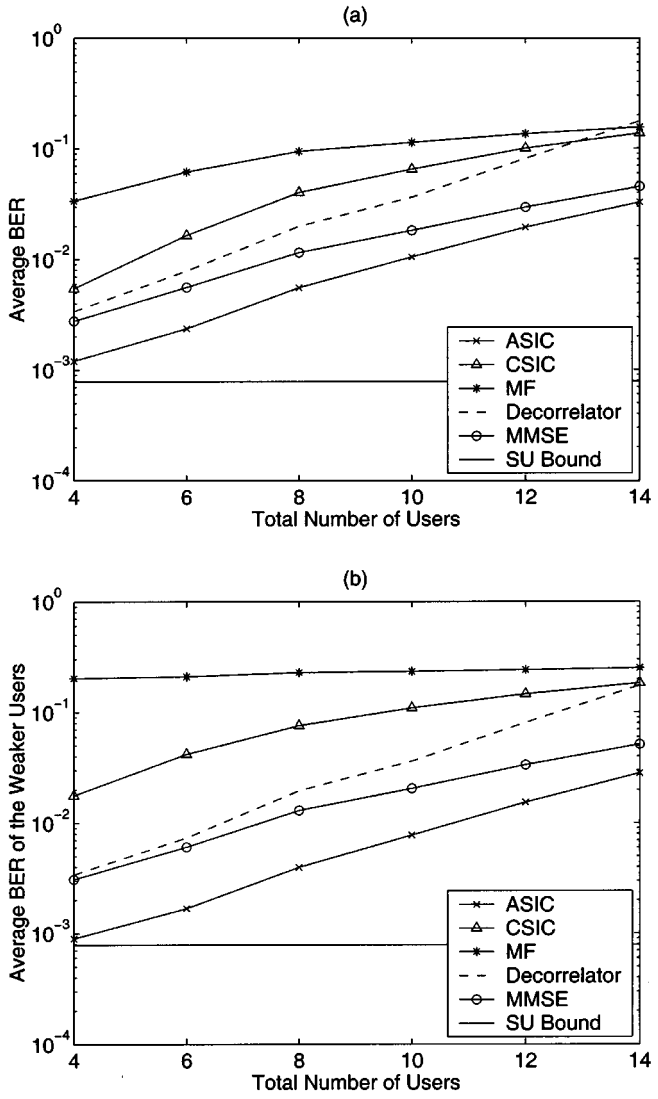

Fig. 11. BER versus system load of the ASIC, CSIC, decorrelator, MMSE receiver, and MF detector for an AWGN channel. (a) Equal power users. (b) Near–far situation.

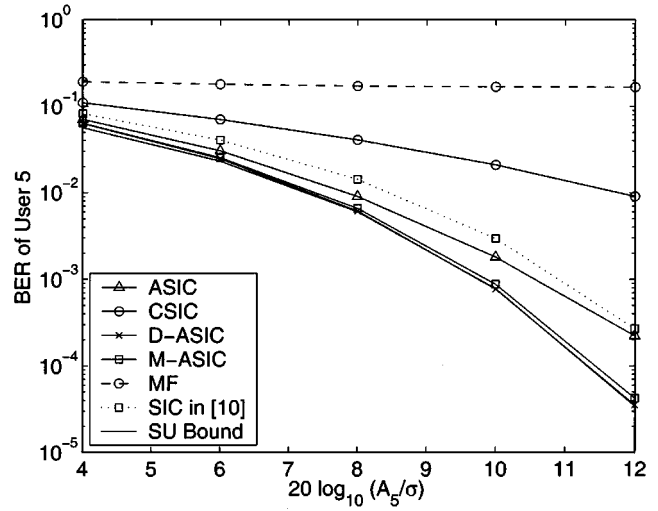

Fig. 12. BER performance of the ASIC with different receiver front ends for an AWGN channel and the same conditions used in Fig. 4.

VII. Conclusion

Using a Wiener model, we have analyzed the steady-state performance of the adaptive successive interference canceler (ASIC), which employs an adaptive interference canceler (AIC) in each stage to minimize the output power. These results indicate that the converged AIC weight, and, thus, the amplitude estimate, are generally biased away from the actual amplitude of the received signal. However, the converged canceler weight is very close to the actual received amplitude of the detected user when the BER of that stage (and of previous stages) is sufficiently small. It was shown that the variance of the amplitude estimate decreases across stages and is controlled by the step-size parameter. In addition, for high SINR conditions, the bias depends primarily on the BER in that stage. Computer simulations demonstrate that the convergence rate is typically fast—a few hundred chips—which is expected since there is only a single adaptive weight in each stage. These examples also illustrate the limitations of the ASIC, in terms of BER performance, due to error-propagation effects (e.g., for high system loads or during deep fades); nevertheless, the ASIC compares favorably with several multiuser receivers (such as the decorrelator and the MMSE receiver) for the scenarios considered here. The corresponding properties of the CSIC were also examined and compared with those of the ASIC. It was shown that for stages with $m > 1$, the amplitude estimate of the CSIC with nonorthogonal spreading sequences is still biased, even when the decisions are always correct. Furthermore, the bias and variance of the CSIC amplitude estimate (as well as the amplitude mismatch) are larger than those of the ASIC, leading to poorer BER performance of the weaker users (especially when many strong users are present).

Appendix A
Derivation of $\hat{A}_{1,o}$

Substituting (4) into (10) yields (recall that $\mathbf{e}_1 = \mathbf{r}$)

$$\hat{A}_{1,o} = A_1 E\Big[b_1\hat{b}_1\Big] + \sum_{j=2}^{K} A_j\rho_{1j}E\Big[b_j\hat{b}_1\Big] + E\Big[\hat{n}_1\hat{b}_1\Big]. \quad \text{(A.1)}$$

In the following sections, we evaluate the three expectations in (A.1) separately.

A. Evaluation of $E[b_1\hat{b}_1]$

Since the product of b_1 and $\hat{b}_1$ can only take on the values ± 1, we can write

$$\begin{aligned} E\Big[b_1\hat{b}_1\Big] &= 1\cdot P\Big(\hat{b}_1 = b_1\Big) + (-1)\cdot P\Big(\hat{b}_1 \neq b_1\Big) \\ &= 1\cdot\Big(1 - P_e^{(1)}\Big) + (-1)\cdot P_e^{(1)} \\ &= 1 - 2P_e^{(1)} \end{aligned} \quad \text{(A.2)}$$

where $P_e^{(1)}$ is the error probability of the first stage. Similarly, $E[b_m\hat{b}_m] = 1 - 2P_e^{(m)}$, where $P_e^{(m)}$ is the error probability of the mth stage.

B. Evaluation of $E[b_j\hat{b}_1]$

Consider the set $B_j \triangleq \{\beta_j = (b_1, \ldots, b_{j-1}, b_{j+1}, \ldots, b_K): b_l = \pm 1, l = 1, \ldots, K, l \neq j\}$. We can write

$$E\left[b_j\hat{b}_1\right] = E_{B_j}E\left[b_j\hat{b}_1|\beta_j\right] = 2^{-(K-1)}\sum_{\beta_j} E\left[b_j\hat{b}_1|\beta_j\right] \tag{A.3}$$

where $\sum_{\beta_j} \triangleq \sum_{b_1}\cdots\sum_{b_{j-1}}\sum_{b_{j+1}}\cdots\sum_{b_K}$ denotes the summations over all possible bit combinations $\{b_l\}$, except for $l = j$. By the definition of conditional expectation

$$E\left[b_j\hat{b}_1|\beta_j\right] = \sum_{b_j}\sum_{\hat{b}_1} b_j\hat{b}_1 P\left(b_j, \hat{b}_1|\beta_j\right) = \tfrac{1}{2}\sum_{b_j}\sum_{\hat{b}_1} b_j\hat{b}_1 P\left(\hat{b}_1|\beta_j, b_j\right) \tag{A.4}$$

where the last equality follows from $P(b_j, \hat{b}_1|\beta_j) = P(\hat{b}_1|\beta_j, b_j)P(\beta_j, b_j)/P(\beta_j) = (1/2)P(\hat{b}_1|\beta_j, b_j)$. Expanding the summations over b_j and $\hat{b}_1$ in (A.4), we have

$$\begin{aligned} &E\left[b_j\hat{b}_1|\beta_j\right] \\ &= \frac{1}{2}\left[\left[Q\left(\frac{-A_j\rho_{1j} - \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right) - \left[1 - Q\left(\frac{-A_j\rho_{1j} - \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right)\right]\right]\right. \\ &\quad - \left[Q\left(\frac{A_j\rho_{1j} - \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right)\right. \\ &\quad \left.\left. - \left[1 - Q\left(\frac{A_j\rho_{1j} - \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right)\right]\right]\right] \end{aligned} \tag{A.5}$$

where the four terms containing Q-functions are due to evaluating $P(\hat{b}_1|\beta_j, b_j)$ for the pairs $(b_j, \hat{b}_1) = (1, 1), (1, -1), (-1, 1)$, and $(-1, -1)$, respectively. Using the identity $Q(-x) - Q(x) = 1 - 2Q(x)$, (A.5) can be simplified to

$$E\left[b_j\hat{b}_1|\beta_j\right] = Q\left(\frac{-A_j\rho_{1j} + \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right) - Q\left(\frac{A_j\rho_{1j} + \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right). \tag{A.6}$$

Thus, inserting (A.6) into (A.3) yields

$$E\left[b_j\hat{b}_1\right] = 2^{-(K-1)}\sum_{\beta_j} Q\left(\frac{-A_j\rho_{1j} + \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right) - Q\left(\frac{A_j\rho_{1j} + \sum_{l\neq j} A_l b_l \rho_{1l}}{\sigma}\right) \tag{A.7}$$

which is the expression used in (11).

To examine the asymptotic properties of $E[b_j\hat{b}_1]$ as $\sigma \to 0$, it is useful to rewrite (A.7) by partially expanding $\sum_{\beta_j}$ over b_1. After exploiting the following symmetry for arbitrary x and y (with all bits taking on the values ± 1)

$$\sum_{b_2}\cdots\sum_{b_{j-1}}\sum_{b_{j+1}}\cdots\sum_{b_K} Q\left(\frac{x + \sum_{l\neq 1,j} A_l b_l \rho_{1l}}{y}\right) = \sum_{b_2}\cdots\sum_{b_{j-1}}\sum_{b_{j+1}}\cdots\sum_{b_K} Q\left(\frac{x - \sum_{l\neq 1,j} A_l b_l \rho_{1l}}{y}\right) \tag{A.8}$$

and substituting $Q(-x) - Q(x) = 1 - 2Q(x)$, (A.7) can be expressed as

$$\begin{aligned} E\left[b_j\hat{b}_1\right] = 2^{-(K-2)}&\sum_{b_2}\cdots\sum_{b_{j-1}}\sum_{b_{j+1}}\cdots\sum_{b_K} \\ &\cdot\left[Q\left(\frac{A_1 - \left(A_j\rho_{1j} - \sum_{l\neq 1,j} A_l b_l \rho_{1l}\right)}{\sigma}\right)\right. \\ &\left. - Q\left(\frac{A_1 + \left(A_j\rho_{1j} - \sum_{l\neq 1,j} A_l b_l \rho_{1l}\right)}{\sigma}\right)\right]. \end{aligned} \tag{A.9}$$

Recall that $Q(x)$ is monotonically decreasing in x with $Q(-\infty) = 1$ and $Q(\infty) = 0$. From (A.9), we see that $E[b_j\hat{b}_1] \to 0$ as $\sigma \to 0$ if and only if the signs of the arguments in both Q-functions agree for all 2^{K-2} possible bit combinations (such that the values of the Q-functions equal 0 or 1 simultaneously), which is equivalent to $A_1 > \sum_{l\neq 1} A_l|\rho_{1l}|$ (i.e., user 1 is sufficiently strong).

C. Evaluation of $E[\hat{n}_1\hat{b}_1]$

Defining the set $B \triangleq \{\beta = (b_1, \ldots, b_K): b_l = \pm 1, l = 1, \ldots, K\}$, it follows that $E[\hat{n}_1\hat{b}_1]$ can be written as

$$E\left[\hat{n}_1\hat{b}_1\right] = E_B E\left[\hat{n}_1\hat{b}_1|\beta\right] = 2^{-K}\sum_{\beta} E\left[\hat{n}_1\hat{b}_1|\beta\right] \tag{A.10}$$

where

$$E\Big[\hat{n}_1\hat{b}_1|\beta\Big] = \int_{-\infty}^{\infty}\zeta\Big[f_{\hat{n}_1,\hat{b}_1|\beta}\Big(\hat{n}_1=\zeta,\,\hat{b}_1=1|\beta\Big) + f_{\hat{n}_1,\hat{b}_1|\beta}\Big(\hat{n}_1=-\zeta,\,\hat{b}_1=-1|\beta\Big)\Big]d\zeta. \quad \text{(A.11)}$$

It is straightforward to see that

$$f_{\hat{n}_1,\hat{b}_1|\beta}\Big(\hat{n}_1=\zeta,\,\hat{b}_1=1|\beta\Big) = f_{\hat{n}_1}\left(\hat{n}_1=\zeta,\,\hat{n}_1>-\sum_{l=1}^{K}A_l b_l\rho_{1l}\right) = \begin{cases}\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\zeta^2/2\sigma^2}, & \zeta>-\displaystyle\sum_{l=1}^{K}A_l b_l\rho_{1l}\\ 0, & \text{otherwise}\end{cases} \quad \text{(A.12)}$$

where the conditioning in the first equality is removed because $\hat{n}_1$ and β are independent. Similarly

$$f_{\hat{n}_1,\hat{b}_1|\beta}\Big(\hat{n}_1=-\zeta,\,\hat{b}_1=-1|\beta\Big) = \begin{cases}\dfrac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\zeta^2/2\sigma^2}, & \zeta>\displaystyle\sum_{l=1}^{K}A_l b_l\rho_{1l}\\ 0, & \text{otherwise.}\end{cases} \quad \text{(A.13)}$$

Substituting (A.12) and (A.13) into (A.11) and integrating yields

$$E\Big[\hat{n}_1\hat{b}_1|\beta\Big] = \sqrt{\frac{2\sigma^2}{\pi}}\exp\left(-\frac{\left(\displaystyle\sum_{l=1}^{K}A_l b_l\rho_{1l}\right)^2}{2\sigma^2}\right) \quad \text{(A.14)}$$

which inserted into (A.10) gives $E[\hat{n}_1\hat{b}_1]$.

APPENDIX B
DERIVATION OF $E[\tilde{A}_m^2]$ AND $\mathrm{var}[\tilde{A}_m]$

Since $\mathrm{var}[\tilde{A}_m] = E[\tilde{A}_m^2] - E^2[\tilde{A}_m]$, and we have already obtained a closed-form expression for $E[\tilde{A}_m]$ in (38), we start with the second moment $E[\tilde{A}_m^2]$. Let B and β be defined as in Appendix A so that we can write

$$E\Big[\tilde{A}_m^2\Big] = E_B E\Big[\tilde{A}_m^2|\beta\Big] = 2^{-K}\sum_{\beta}E\Big[\tilde{A}_m^2|\beta\Big]. \quad \text{(B.1)}$$

In order to evaluate $E[\tilde{A}_m^2|\beta]$, we employ the following useful result (which can easily be shown by using a transformation of random variables [29]): If X is a Gaussian random variable with mean m_x and variance σ_x^2, then the pdf of $Y=|X|$ is

$$f_Y(y) = \begin{cases}\dfrac{1}{\sqrt{2\pi\sigma_x^2}}\Big[e^{-(y-m_x)^2/2\sigma_x^2} + e^{-(y+m_x)^2/2\sigma_x^2}\Big], & y\ge 0\\ 0, & y<0.\end{cases} \quad \text{(B.2)}$$

Recall that

$$\tilde{A}_m = \Big|\mathbf{s}_m^T\tilde{\mathbf{e}}_m\Big| = \left|\sum_{l=1}^{K}A_l b_l\tilde{\rho}_{ml}+\tilde{n}_m\right| \triangleq |\tilde{z}_m|. \quad \text{(B.3)}$$

Note that given $\beta=(b_1,\ldots,b_K)$, $\tilde{z}_m$ is a Gaussian random variable with mean $\sum_{l=1}^{K}A_l b_l\tilde{\rho}_{ml}$ and variance $\tilde{\sigma}_m^2$. Hence, given β, $\tilde{A}_m$ has the pdf $f_{\tilde{A}_m|\beta}(\zeta)$ with the form of (B.2), except that m_x and σ_x are replaced by $\sum_{l=1}^{K}A_l b_l\tilde{\rho}_{ml}$ and $\tilde{\sigma}_m$, respectively. It immediately follows from $E[Y^2]=m_x^2+\sigma_x^2$ that

$$E\Big[\tilde{A}_m^2|\beta\Big] = \left(\sum_{l=1}^{K}A_l b_l\tilde{\rho}_{ml}\right)^2+\tilde{\sigma}_m^2. \quad \text{(B.4)}$$

Thus, inserting (B.4) into (B.1) and using $\sum_{b_j}\sum_{b_l}b_j b_l=\delta_{jl}$ (the Kronecker delta function) yields (39), from which (40) immediately follows.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful suggestions and comments and Dr. Y. Cho for discussions of the SIC in [10]. They would also like to thank one of the reviewers for pointing out that the ASIC/CSIC analyses can be generalized to arbitrary linear filter front ends.

REFERENCES

[1] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
[2] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Reading, MA: Addison-Wesley, 1995.
[3] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.
[4] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, vol. 34, pp. 124–136, Oct. 1996.
[5] 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.
[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. 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.
[8] ——, "Performance comparison of a DS/CDMA system using a successive interference cancellation (IC) scheme and a parallel IC scheme under fading," in *Proc. IEEE Int. Conf. Commun.*, New Orleans, LA, May 1994, pp. 510–514.
[9] D. S. Chen and S. Roy, "An adaptive multiuser receiver for CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 808–816, June 1994.
[10] Y. Cho and J. H. Lee, "Analysis of an adaptive SIC for near–far resistant DS-CDMA," *IEEE Trans. Commun.*, vol. 46, pp. 1429–1432, Nov. 1998.
[11] B. Zhu, N. Ansari, and Z. Siveski, "Convergence and stability analysis of a synchronous adaptive CDMA receiver," *IEEE Trans. Commun.*, vol. 43, pp. 3073–3079, Dec. 1995.
[12] S. J. Baines, A. G. Burr, and T. C. Tozer, "Adaptive architecture for signal separation and interference suppression in DS-CDMA systems," *IEE Electron. Lett.*, vol. 32, pp. 1057–1058, June 1996.
[13] G. Xue, J. Weng, T. Le-Ngoc, and S. Tahar, "Adaptive multistage parallel interference cancellation for CDMA," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1815–1827, Oct. 1999.
[14] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[15] K.-C. Lai, J. J. Shynk, M. Motamed, and R. P. Gooch, "Adaptive successive interference cancellation for the IS-95 uplink," in *Proc. IEEE Vech. Technol. Conf.*, Boston, MA, Sept. 2000, pp. 1187–1192.

[16] K.-C. Lai, R. E. Cagley, J. J. Shynk, M. Motamed, and R. P. Gooch, "Comparative performance of adaptive receivers for demodulating IS-95 downlink data," in *Proc. IEEE Adaptive Syst. Signal Process., Commun., Contr. Symp.*, Banff, AB, Canada, Oct. 2000, pp. 414–419.
[17] Electron. Ind. Assoc., "Mobile station–base station compatibility standard for dual-mode wideband spread spectrum cellular system," EIA/TIA IS-95, Washington, DC, July 1993.
[18] L. K. Rasmussen, T. J. Lim, and A.-L. Johansson, "A matrix-algebraic approach to successive interference cancellation in CDMA," *IEEE Trans. Commun.*, vol. 48, pp. 145–151, Jan. 2000.
[19] J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.
[20] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*. Boston, MA: Kluwer, 1984.
[21] R. Lupas and S. Verdú, "Linear multi-user detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.
[22] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimal detectors for coherent multi-user communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.
[23] K.-C. Lai and J. J. Shynk, "On the performance of the successive interference canceler for DS/CDMA signals," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 1999, pp. 1594–1600.
[24] ——, "Error-rate analysis of the adaptive successive interference canceler for DS/CDMA signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, June 2000, pp. 2845–2848.
[25] D. R. Brown and C. R. Johnson, Jr., "SINR, power efficiency, and theoretical system capacity of parallel interference cancellation," in *Proc. Thirty-Fourth Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 2000, pp. TA2.1–TA2.6.
[26] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.
[27] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. Orlando, FL: Academic, 1985.
[28] W. C. Jakes, Ed., *Microwave Mobile Communications*. Piscataway, NJ: IEEE, 1994.
[29] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

Kuei-Chiang Lai received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1993 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 1997 and 2001, respectively.

His research interests are in the area of signal processing for communications.

John J. Shynk (SM'91) received the B.S. degree in systems engineering from Boston University, Boston, MA, in 1979 and the M.S. degree in electrical engineering and in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1980, 1985, and 1987, respectively.

From 1979 to 1982, he was a Member of Technical Staff with the Data Communications Performance Group, AT&T Bell Laboratories, Holmdel, NJ, where he formulated performance models for voiceband data communications. He was a Research Assistant from 1982 to 1986 with the Department of Electrical Engineering, Stanford University, where he worked on frequency-domain implementations of adaptive IIR filter algorithms. From 1985 to 1986, he was also an Instructor at Stanford University, teaching courses on digital signal processing and adaptive systems. Since 1987, he has been with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, where he is currently a Professor. His research interests include adaptive signal processing, adaptive beamforming, wireless communications, and direction-of-arrival estimation. He has served as Editor for adaptive signal processing for the *International Journal of Adaptive Control and Signal Processing*.

Dr. Shynk served as an Associate Editor for adaptive filtering for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He was Technical Program Chair of the 1992 International Joint Conference on Neural Networks. He is a Member of the IEEE Signal Processing for Communications Technical Committee.