Adaptive Bayesian Multiuser Detection for Synchronous CDMA with Gaussian and Impulsive Noise

Xiaodong Wang, *Member, IEEE,* and Rong Chen

Abstract—We consider the problem of simultaneous parameter estimation and data restoration in a synchronous CDMA system in the presence of either additive Gaussian or additive impulsive white noise with unknown parameters. The impulsive noise is modeled by a two-term Gaussian mixture distribution. Bayesian inference of all unknown quantities is made from the superimposed and noisy received signals. The Gibbs sampler (a Markov chain Monte Carlo procedure) is employed to calculate the Bayesian estimates. The basic idea is to generate ergodic random samples from the joint posterior distribution of all unknown and then to average the appropriate samples to obtain the estimates of the unknown quantities. Adaptive Bayesian multiuser detectors based on the Gibbs sampler are derived for both the Gaussian noise synchronous CDMA channel and the impulsive noise synchronous CDMA channel. A salient feature of the proposed adaptive Bayesian multiuser detectors is that they can incorporate the *a priori* symbol probabilities, and they produce as output the *a posteriori* symbol probabilities. (That is, they are "soft-input soft-output" algorithms.) Hence, these methods are well suited for iterative processing in a coded system, which allows the adaptive Bayesian multiuser detector to refine its processing based on the information from the decoding stage, and vice versa—a receiver structure termed *adaptive Turbo multiuser detector.*

Index Terms—Adaptive multiuser detection, Bayesian inference, Gibbs sampler, impulsive noise, iterative processing.

I. INTRODUCTION

THE THEME of this paper is to treat three related problems in multiuser detection under a general Bayesian framework. These problems are

i) optimal multiuser detection in the presence of unknown channel parameters;

ii) optimal multiuser detection in impulsive ambient noise;

iii) multiuser detection for coded CDMA systems.

We first provide a perspective on the related works in these three areas.

Optimal Multiuser Detection with Unknown Parameters: The optimal multiuser detection algorithms with known

X. Wang is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: wangx@ee.tamu.edu).

R. Chen is with the Department of Statistics, Texas A&M University, College Station, TX 77843 USA.

channel parameters, that is, the multiuser maximum-likelihood sequence detector (MLSD) and the multiuser maximum *a posteriori* symbol probability (MAP) detector, were first investigated in [46] and [47] (cf. [49]). The analysis of the computational complexity and the proof that the optimal multiuser detection problem is combinatorially hard appeared in [46] and [48]. When the channel parameters (e.g., the received amplitudes and the noise variance) are unknown, it is of interest to study the problem of joint multiuser channel parameter estimation and data detection from the received waveform. This problem was first treated in [36], where a solution based on the expectation-maximization (EM) algorithm is derived. In [42], the problem of sequential multiuser amplitude estimation in the presence of unknown data is studied, and an approach based on stochastic approximation is proposed. In [55], a tree-search algorithm is given for joint data detection and amplitude estimation. Other works concerning multiuser detection with unknown channel parameters include [14], [23], [24], [33], [35], and [43].

Multiuser Detection with Impulsive Noise: To date, most of the work on multiuser detection assumes that the channel ambient noise is Gaussian. However, in many physical channels where multiuser detection may be applied, such as urban and indoor radio channels [8], [9], [28], [29], [31], [32] and underwater acoustic channels [11], [30], the ambient noise is known through experimental measurements to be decidedly non-Gaussian, due to the impulsive nature of the man-made electromagnetic interference and a great deal of natural noise as well. The results of an early study of error rates in non-Gaussian CDMA channels are found in [1]–[3], in which the performance of the conventional and modified conventional detectors is shown to depend significantly on the shape of the ambient noise distribution. In [37], it is observed that the performance gains afforded by maximum likelihood multiuser detection in impulsive noise can be substantial when compared with optimum multiuser detection based on a Gaussian noise assumption. In [52], robust multiuser detection methods for impulsive noise CDMA channels based on the Huber robust regression technique are proposed.

Multiuser Detection for Coded CDMA: Most CDMA systems employ error control coding to protect the transmitted data from being corrupted by the channel. Some recent work has addressed multiuser detection for coded CDMA systems. In [18], the optimal decoding scheme for convolutionally coded CDMA system is studied, which is shown to have a prohibitive computational complexity. In [19], some low-complexity receivers
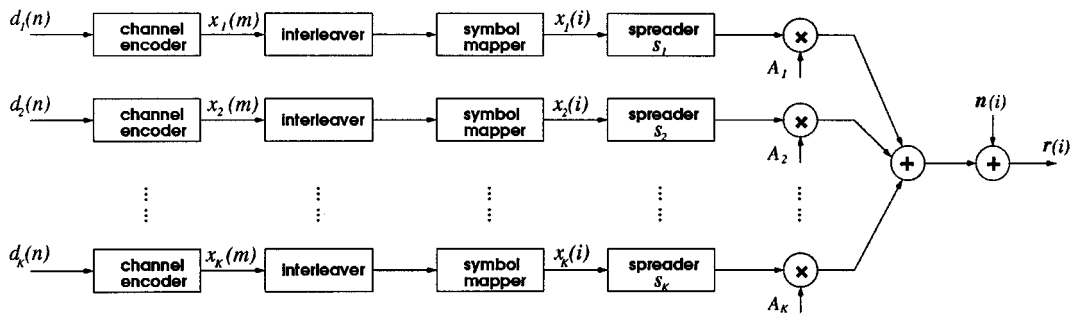
Fig. 1. Coded synchronous CDMA communication system.

that perform multiuser symbol detection and decoding either separately or jointly are studied. In [34], [38], and [51], Turbo multiuser detection schemes for coded CDMA systems are proposed, which iterate between multiuser detection and channel decoding to successively improve the receiver performance.

In this paper, we present novel adaptive Bayesian multiuser detection techniques for synchronous CDMA communications with unknown channel parameters in both Gaussian and impulsive ambient noise channels. The impulsive noise is modeled by a two-term Gaussian mixture distribution. We consider Bayesian inference of all unknown quantities (e.g., the received amplitudes, the data symbols, and the noise parameters) from the received waveforms. A Markov chain Monte Carlo procedure called the Gibbs sampler is employed to calculate the Bayesian estimates. The performance of the proposed adaptive multiuser detectors is demonstrated via simulations. The proposed Bayesian multiuser detectors can naturally exploit the structure of the coded signals. Another salient feature of the proposed methods is that being soft-input soft-output demodulation algorithms, they can be used in conjunction with soft channel decoding algorithm to accomplish iterative joint adaptive multiuser detection and decoding—the so-called *adaptive Turbo multiuser detection.*

The rest of the paper is organized as follows. In Section II, the system under study is described. In Section III, some background material on the Gibbs sampler is provided. The problems of adaptive Bayesian multiuser detection in Gaussian noise and impulsive noise synchronous CDMA channels are treated in Sections IV and V, respectively. In Section VI, an adaptive Turbo multiuser detection scheme is presented. Some discussions, including a decoder-assisted convergence assessment scheme and a code-constrained Bayesian multiuser detector, are found in Section VII. Simulation results are provided in Section VIII. Finally, Section IX contains the conclusions.

II. SYSTEM DESCRIPTION

We consider a coded synchronous CDMA system with K users, employing normalized modulation waveforms $s_1, s_2, \cdots, s_K$ and signaling through a channel with additive white noise. The block diagram of the transmitter end of such a system is shown in Fig. 1. The binary information bits $\{d_k(n)\}$ for user k are encoded using some channel code (e.g., block code, convolutional code, or Turbo code), resulting in a code bit stream $\{x_k(m)\}$. A code-bit interleaver is used to reduce

the influence of the error bursts at the input of the channel decoder. The interleaved code bits are then mapped to BPSK symbols, yielding symbol stream $\{x_k(i)\}$. Each data symbol is then modulated by a spreading waveform s_k and transmitted through the channel. The received signal is the superposition of the K users' transmitted signals plus the ambient noise, which is given by

$$r(i) = \sum_{k=1}^{K} A_k x_k(i) s_k + n(i), \qquad i = 0, \cdots, M-1. \quad (1)$$

In (1), M is the number of data symbols per user per frame; A_k, $x_k(i)$, and s_k denote, respectively, the amplitude, the ith symbol, and the normalized spreading waveform of the kth user; and $n(i) = [n_0(i) \; n_1(i) \; \cdots \; n_{P-1}(i)]^T$ is a zero-mean white noise vector. The spreading waveform is of the form

$$s_k = \frac{1}{\sqrt{P}} [\beta_{k,0} \; \beta_{k,1} \; \cdots \; \beta_{k,P-1}]^T$$
$$\beta_{k,j} \in \{+1, -1\} \quad (2)$$

where P is the spreading factor. It is assumed that the receiver knows the spreading waveforms of all active users in the system. Define the following *a priori* symbol probabilities:

$$\rho_k(i) \overset{\Delta}{=} P[x_k(i) = +1], \qquad i = 0, \cdots, M-1$$
$$k = 1, \cdots, K. \quad (3)$$

Note that when no prior information is available, then $\rho_k(i) = 1/2$, i.e., all symbols are equally likely.

It is further assumed that the additive ambient channel noise vector $\{n(i)\}$ is a sequence of zero-mean independent and identically distributed (i.i.d.) random vectors, and it is independent of the symbol sequences $\{x_k(i)\}_{k=1}^{K}$. Moreover, the noise vector $n(i)$ is assumed to consist of i.i.d. samples $\{n_j(i)\}_{j=0}^{P-1}$. In this paper, we consider two types of noise distributions corresponding to the additive Gaussian noise and the additive impulsive noise, respectively. For the former case, the noise $n_j(i)$ is assumed to have a Gaussian distribution, i.e.,

$$n_j(i) \sim \mathcal{N}(0, \sigma^2) \quad (4)$$

where σ^2 is the variance of the noise. For the latter case, the noise $n_j(i)$ is assumed to have a two-term Gaussian mixture distribution, i.e.,

$$n_j(i) \sim (1-\epsilon)\mathcal{N}(0, \sigma_1^2) + \epsilon\mathcal{N}(0, \sigma_2^2) \quad (5)$$

with $0 < \epsilon < 1$ and $\sigma_1^2 < \sigma_2^2$. Here, the term $\mathcal{N}(0, \sigma_1^2)$ represents the nominal ambient noise, and the term $\mathcal{N}(0, \sigma_2^2)$ represents an impulsive component, with ϵ representing the probability that an impulse occurs. The total noise variance under distribution (5) is given by

$$\sigma^2 = (1 - \epsilon)\sigma_1^2 + \epsilon\sigma_2^2. \tag{6}$$

Denote $Y \triangleq \{r(0), r(1), \cdots, r(M-1)\}$. In Sections IV and V, we consider the problem of estimating the *a posteriori* probabilities of the transmitted symbols

$$P[x_k(i) = +1 \,|\, Y], \qquad i = 0, \cdots, M-1; \; k = 1, \cdots, K \tag{7}$$

based on the received signals Y and the prior information $\{\rho_k(i)\}_{k=1; i=0}^{K; M-1}$, without knowing the channel amplitudes $\{A_k\}_{k=1}^{K}$ and the noise parameters (i.e., σ^2 for Gaussian noise and ϵ, σ_1^2, and σ_2^2 for impulsive noise). These *a posteriori* probabilities are then used by the channel decoder to decode the information bits $\{d_k(n)\}$ shown in Fig. 1, which will be discussed in Section VI.

III. GIBBS SAMPLER

Over the last decade or so, a large body of methods has emerged based on iterative Monte Carlo techniques that are especially useful in computing Bayesian solutions to estimation problems with high parameter dimensions. These methods are based on the theory of Markov chain limiting behavior and are collectively known as *Markov chain Monte Carlo* (MCMC) techniques [44]. Most of these methods are aimed at estimating the entire posterior density and not just finding the maximum *a posteriori* (MAP) estimates of the parameters. One of the most popular of these methods is known as the *Gibbs sampler* [15], which is described next.

Let $\theta = [\theta_1 \; \cdots \; \theta_d]^T$ be a vector of unknown parameters, and let Y be the observed data. Suppose that we are interested in finding the *a posteriori* marginal distribution of some parameter, say, θ_j, conditioned on the observation Y, i.e., $p(\theta_j \,|\, Y)$, $1 \leq j \leq d$. Direct evaluation involves integrating out the rest of the parameters from the joint *a posteriori* density, i.e.,

$$p(\theta_j \,|\, Y) = \iint \cdots \int p(\theta \,|\, Y) \, d\theta_1 \cdots d\theta_{j-1} \, d\theta_{j+1} \cdots d\theta_d. \tag{8}$$

In most cases, such a direct evaluation is computationally infeasible, especially when the parameter dimension d is large. The Gibbs sampler is a Monte Carlo procedure for numerical evaluation of the above multidimensional integral. The basic idea is to generate random samples from the joint posterior distribution $p(\theta \,|\, Y)$ and then to estimate any marginal distribution using these samples. Given the initial values $\theta^{(0)} = [\theta_1^{(0)} \; \cdots \; \theta_d^{(0)}]^T$, this algorithm iterates the following loop:

- Draw sample $\theta_1^{(n+1)}$ from $p(\theta_1 \,|\, \theta_2^{(n)}, \cdots, \theta_d^{(n)}, Y)$.
- Draw sample $\theta_2^{(n+1)}$ from $p(\theta_2 \,|\, \theta_1^{(n+1)}, \theta_3^{(n)}, \cdots, \theta_d^{(n)}, Y)$.

$\qquad \vdots$

- Draw sample $\theta_d^{(n+1)}$ from $p(\theta_d \,|\, \theta_1^{(n+1)}, \cdots, \theta_{d-1}^{(n+1)}, Y)$.

Under regularity conditions, in the steady state, the sequence of sample vectors $\cdots, \theta^{(n-1)}, \theta^{(n)}, \theta^{(n+1)}, \cdots$ is a realization of a homogeneous Markov chain with the transition kernel from state θ' to state θ given by

$$K(\theta', \theta) = p(\theta_1 \,|\, \theta_2', \cdots, \theta_d', Y) \; p(\theta_2 \,|\, \theta_1, \theta_3', \cdots, \theta_d', Y) \\ \cdots \; p(\theta_d \,|\, \theta_1, \cdots, \theta_{d-1}, Y). \tag{9}$$

The convergence behavior of the Gibbs sampler is investigated in [13], [15], [17], [26], [41], and [45], and general conditions are given for the following two results:

- The distribution of $\theta^{(n)}$ converges geometrically to $p(\theta \,|\, Y)$, as $n \longrightarrow \infty$.
- $\frac{1}{N} \sum_{n=1}^{N} f\left(\theta^{(n)}\right) \xrightarrow{\text{a.s.}} \int f(\theta) p(\theta \,|\, Y) \, d\theta$, as $N \longrightarrow \infty$, for any integrable function f.

The Gibbs sampler requires an initial transient period to converge to equilibrium. The initial period of length n_0 is known as the "burn-in" period, and the first n_0 samples should always be discarded. Detecting convergence is usually done in some *ad hoc* way. Some methods are found in [44].

The roots of MCMC methods can be traced back to the well-known Metropolis algorithm [27], which was initially used to investigate the equilibrium properties of molecules in a gas. The first use of the Metropolis algorithm in a statistical context is found in [22]. The Gibbs sampler, which is a special case of the Metropolis algorithm, was so termed in the seminal paper [17] on image processing. It is brought to statistical prominence by [15], where it was observed that many Bayesian computation could be carried out via the Gibbs sampler. For tutorials on the Gibbs sampler, see [4] and [12].

IV. ADAPTIVE BAYESIAN MULTIUSER DETECTION IN GAUSSIAN NOISE

In this section, we consider the problem of computing the *a posteriori* probabilities in (7) under the assumption that the ambient noise distribution is Gaussian. That is, the pdf of $n(i)$ in (1) is given by

$$p(n(i)) = \frac{1}{(2\pi\sigma^2)^{P/2}} \exp\left(-\frac{\|n(i)\|^2}{2\sigma^2}\right). \tag{10}$$

Denote

$$x(i) \triangleq [x_1(i) \quad x_2(i) \quad \cdots \quad x_K(i)]^T$$
$$i = 0, 1, \cdots, M-1$$
$$B(i) \triangleq \mathtt{diag}(x_1(i), x_2(i), \cdots, x_K(i))$$
$$i = 0, 1, \cdots, M-1$$
$$X \triangleq [x(0) \quad x(1) \quad \cdots \quad x(M-1)]$$
$$Y \triangleq [r(0) \quad r(1) \quad \cdots \quad r(M-1)]$$
$$a \triangleq [A_1 \quad A_2 \quad \cdots \quad A_K]^T$$
$$A \triangleq \mathtt{diag}(A_1, A_2, \cdots, A_K)$$
$$S \triangleq [s_1 \quad s_2 \quad \cdots \quad s_K].$$

Then, (1) can be written as

$$r(i) = SAx(i) + n(i) \qquad (11)$$
$$= SB(i)a + n(i), \qquad i = 0, 1, \cdots, M-1. \quad (12)$$

The problem is solved under a Bayesian framework. First, the unknown quantities a, σ^2, and X are regarded as realizations of random variables with some prior distributions. The Gibbs sampler, which is a Monte Carlo method, is then employed to calculate the maximum *a posteriori* (MAP) estimates of these unknowns.

A. Bayesian Inference

Assume that the unknown quantities a, σ^2 and X are independent of each other and have prior distributions $p(a)$, $p(\sigma^2)$, and $p(X)$, respectively. Since $\{n(i)\}_{i=0}^{M-1}$ is a sequence of independent Gaussian vectors, using (10) and (11), the joint posterior distribution of these unknown quantities (a, σ^2, X) based on the received signal Y takes the form of

$$
\begin{aligned}
&p(a, \sigma^2, X \,|\, Y)\\
&\quad = p(Y \,|\, a, \sigma^2, X)\, p(a) p(\sigma^2)\, p(X) / p(Y)\\
&\quad = C \cdot \left(\frac{1}{\sigma^2}\right)^{PM/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \|r(i) - SAx(i)\|^2\right)\\
&\qquad \cdot p(a) p(\sigma^2)\, p(X)
\end{aligned}
\qquad (13)
$$

where C is a normalization constant independent of the unknown parameters (a, σ^2, X). The *a posteriori* probabilities (7) of the transmitted symbols can then be calculated from the joint posterior distribution (13) according to

$$
\begin{aligned}
&P[x_k(i) = +1 \,|\, Y]\\
&\quad = \sum_{X: x_k(i)=+1} p(X \,|\, Y)\\
&\quad = \sum_{X: x_k(i)=+1} \int p(a, \sigma^2, X \,|\, Y)\, da\, d\sigma^2.
\end{aligned}
\qquad (14)
$$

Clearly, the computation in (14) involves 2^{KM-1} multidimensional integrals, which is certainly infeasible for any practical implementations. To avoid the direct evaluation of the Bayesian estimate (14), we resort to the Gibbs sampler discussed in Section III. The basic idea is to generate ergodic random samples $\{a^{(n)}, \sigma^{2(n)}, X^{(n)}: n = n_0, n_0 + 1, \cdots\}$ from the posterior distribution (13), and then to average $\{x_k(i)^{(n)}: n = n_0, n_0+1, \cdots\}$ to obtain an approximation of the *a posteriori* probabilities in (14).

B. Prior Distributions

General Considerations:

Noninformative Priors: In Bayesian analysis, prior distributions are used to incorporate the prior knowledge about the unknown parameters. When such prior knowledge is limited, the prior distributions should be chosen such that they have a minimal impact on the posterior distribution. Such priors are termed *noninformative*. The rationale for using noninformative prior distributions is to "let the data speak for themselves," so that inferences are unaffected by information external to current data [10], [16].

Conjugate Priors: Another consideration in the selection of the prior distributions is to simplify computations. To that end, *conjugate priors* are usually used to obtain simple analytical forms for the resulting posterior distributions. The property that the posterior distribution belongs to the same distribution family as the prior distribution is called conjugacy. The conjugate family of distributions is mathematically convenient in that the posterior distribution follows a known parametric form [10], [16]. Finally, to make the Gibbs sampler more computationally efficient, the priors should also be chosen such that the conditional posterior distributions are easy to simulate.

For an introductory treatment of the Bayesian philosophy, including the selection of prior distributions, see [10], [16], and [25]. An account of criticism of the Bayesian approach to data analysis can be found in [5] and [40], and a defense of "The Bayesian Choice" can be found in [39].

Prior Distributions of the Unknowns: Following the general guidelines in Bayesian analysis [10], [16], [25], we choose the conjugate prior distributions for the unknown parameters $p(a)$, $p(\sigma^2)$ and $p(X)$, as follows.

For the unknown amplitude vector a, a truncated Gaussian prior distribution is assumed

$$p(a) \propto \mathcal{N}(a_0, \Sigma_0) I_{\{a > 0\}} \qquad (15)$$

where $I_{\{a > 0\}}$ is an indicator that is 1 if all elements of a are positive and it is zero otherwise. Note that large value of Σ_0 corresponds to the less-informative prior. Note also that although user amplitudes are assumed to be independent, the matrix Σ_0 is not restricted to be diagonal. For the noise variance σ^2, an inverse chi-square prior distribution is assumed

$$
\begin{aligned}
p(\sigma^2) &= \frac{\left(\dfrac{\nu_0 \lambda_0}{2}\right)^{\nu_0/2}}{\Gamma\left(\dfrac{\nu_0}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{(\nu_0/2)+1} \exp\left(-\frac{\nu_0 \lambda_0}{2\sigma^2}\right)\\
&\sim \chi^{-2}(\nu_0, \lambda_0),
\end{aligned}
\qquad (16)
$$

or

$$\frac{\nu_0 \lambda_0}{\sigma^2} \sim \chi^2(\nu_0). \qquad (17)$$

The small value of ν_0 corresponds to the less informative priors (roughly the prior knowledge is worth ν_0 data points). The value of $\nu_0 \lambda_0$ reflects the prior belief of the value of σ^2. Finally, since the symbols $\{x_k(i)\}_{k=1; i=0}^{K; M-1}$ are assumed to be independent, the prior distribution $p(X)$ can be expressed in terms of the prior symbol probabilities defined in (3) as

$$p(X) = \prod_{i=0}^{M-1} \prod_{k=1}^{K} \rho_k(i)^{\delta_{ki}} [1 - \rho_k(i)]^{1-\delta_{ki}} \qquad (18)$$

where δ_{ki} is the indicator such that $\delta_{ki} = 1$ if $x_k(i) = +1$ and $\delta_{ki} = 0$ if $x_k(i) = -1$.

C. Conditional Posterior Distributions

The following conditional posterior distributions are required by the Gibbs multiuser detector in Gaussian noise. The derivations are found in Appendix A.

1) The conditional distribution of the amplitude vector $\boldsymbol{a}$, given σ^2, $\boldsymbol{X}$ and $\boldsymbol{Y}$, is given by

$$p(\boldsymbol{a}\,|\,\sigma^2,\,\boldsymbol{X},\,\boldsymbol{Y}) \propto \mathcal{N}(\boldsymbol{a}_*,\,\boldsymbol{\Sigma}_*)I_{\{\boldsymbol{a}>\boldsymbol{0}\}} \qquad (19)$$

with

$$\boldsymbol{\Sigma}_*^{-1} \triangleq \boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2}\sum_{i=0}^{M-1} \boldsymbol{B}(i)\boldsymbol{R}\boldsymbol{B}(i), \qquad (20)$$

and

$$\boldsymbol{a}_* \triangleq \boldsymbol{\Sigma}_*\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{a}_0 + \frac{1}{\sigma^2}\sum_{i=0}^{M-1} \boldsymbol{B}(i)\boldsymbol{S}^T\boldsymbol{r}(i)\right) \qquad (21)$$

where in (20) $\boldsymbol{R} \triangleq \boldsymbol{S}^T\boldsymbol{S}$.

2) The conditional distribution of the noise variance σ^2 given $\boldsymbol{a}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$p(\sigma^2\,|\,\boldsymbol{a},\,\boldsymbol{X},\,\boldsymbol{Y}) \sim \chi^{-2}\left(\nu_0 + PM,\,\frac{\nu_0\lambda_0 + s^2}{\nu_0 + PM}\right) \qquad (22)$$

or

$$\frac{(\nu_0\lambda_0 + s^2)}{\sigma^2} \sim \chi^2(\nu_0 + PM) \qquad (23)$$

with

$$s^2 \triangleq \sum_{i=0}^{M-1} \|\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(i)\|^2. \qquad (24)$$

3) The conditional probabilities of $x_k(i) = \pm 1$, given $\boldsymbol{a}$, σ^2, $\boldsymbol{X}_{ki}$, and $\boldsymbol{Y}$, can be obtained from [where $\boldsymbol{X}_{ki}$ denotes the set containing all elements of $\boldsymbol{X}$ except for $x_k(i)$]

$$\frac{P[x_k(i) = +1\,|\,\boldsymbol{a},\,\sigma^2,\,\boldsymbol{X}_{ki},\,\boldsymbol{Y}]}{P[x_k(i) = -1\,|\,\boldsymbol{a},\,\sigma^2,\,\boldsymbol{X}_{ki},\,\boldsymbol{Y}]}$$
$$= \frac{\rho_k(i)}{1 - \rho_k(i)}\cdot\exp\left\{\frac{2A_k}{\sigma^2}\,\boldsymbol{s}_k^T\left[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}_k^0(i)\right]\right\}$$
$$k = 1,\cdots,K;\; i = 0,\cdots,M-1 \qquad (25)$$

where

$$\boldsymbol{x}_k^0(i) \triangleq [x_1(i),\cdots,x_{k-1}(i),\,0,\,x_{k+1}(i),\cdots,x_K(i)]^T.$$

D. Gibbs Multiuser Detector in Gaussian Noise

Using the above conditional posterior distributions, the Gibbs sampling implementation of the adaptive Bayesian multiuser detector in Gaussian noise proceeds iteratively as follows. Given the initial values of the unknown quantities $\{\boldsymbol{a}^{(0)},\,\sigma^{2^{(0)}},\,\boldsymbol{X}^{(0)}\}$ drawn from their prior distributions, and for $n = 1, 2, \cdots$, we have the following.

1) Draw $\boldsymbol{a}^{(n)}$ from $p(\boldsymbol{a}\,|\,\sigma^{2^{(n-1)}},\,\boldsymbol{X}^{(n-1)},\,\boldsymbol{Y})$ given by (19).
2) Draw $\sigma^{2^{(n)}}$ from $p(\sigma^2\,|\,\boldsymbol{a}^{(n)},\,\boldsymbol{X}^{(n-1)},\,\boldsymbol{Y})$ given by (23).

3) For $i = 0, 1, \cdots, M - 1$
   For $k = 1, 2, \cdots, K$
   Draw $x_k(i)^{(n)}$ from $P[x_k(i)\,|\,\boldsymbol{a}^{(n)},\,\sigma^{2^{(n)}},\,\boldsymbol{X}_{ki}^{(n)},\,\boldsymbol{Y}]$ given by (25)
   where

$$\boldsymbol{X}_{ki}^{(n)} \triangleq \Big\{\boldsymbol{x}(0)^{(n)},\cdots,\boldsymbol{x}(i-1)^{(n)},\,x_1(i)^{(n)},\cdots$$
$$x_{k-1}(i)^{(n)},\,x_{k+1}(i)^{(n-1)},\cdots,x_K(i)^{(n-1)}$$
$$\boldsymbol{x}(i+1)^{(n-1)},\cdots,\boldsymbol{x}(M-1)^{(n-1)}\Big\}.$$

Note that to draw samples of $\boldsymbol{a}$ from (15) or (19), the so-called *rejection method* [50] can be used. For instance, after a sample is drawn from $\mathcal{N}(\boldsymbol{a}_0,\,\boldsymbol{\Sigma}_0)$ or $\mathcal{N}(\boldsymbol{a}_*,\,\boldsymbol{\Sigma}_*)$, check to see if the constraint $A_k > 0$, $k = 1,\cdots,K$ is satisfied; if not, the sample is rejected, and a new sample is drawn from the same distribution. The procedure continues until a sample is obtained that satisfies the constraint.

To ensure convergence, the above procedure is usually carried out for $(n_0 + N)$ iterations, and samples from the last N iterations are used to calculate the Bayesian estimates of the unknown quantities. In particular, the *a posteriori* symbol probabilities in (14) are approximated as

$$\hat{P}[x_k(i) = +1\,|\,\boldsymbol{Y}] \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N} \delta_{ki}^{(n)} \qquad (26)$$

where $\delta_{ki}^{(n)}$ is the indicator such that $\delta_{ki}^{(n)} = 1$ if $x_k^{(n)} = +1$ and $\delta_{ki}^{(n)} = 0$ if $x_k^{(n)} = -1$. An MAP decision on the symbol $x_k(i)$ is then given by

$$\widehat{x_k(i)} = \arg\max_{b\in\{+1,-1\}} \hat{P}[x_k(i) = b\,|\,\boldsymbol{Y}]. \qquad (27)$$

Furthermore, if desired, the estimates of the amplitude vector $\boldsymbol{a}$ and the noise variance σ^2 can also be obtained from the corresponding sample means

$$E\{\boldsymbol{a}\,|\,\boldsymbol{Y}\} \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N} \boldsymbol{a}^{(n)} \qquad (28)$$

and

$$E\{\sigma^2\,|\,\boldsymbol{Y}\} \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N} \sigma^{2^{(n)}}. \qquad (29)$$

The posterior variances of $\boldsymbol{a}$ and σ^2, which reflect the uncertainty in estimating these quantities on the basis of $\boldsymbol{Y}$, can also be approximated by the sample variances, as

$$\text{Cov}\{\boldsymbol{a}\,|\,\boldsymbol{Y}\} \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N} \left[\boldsymbol{a}^{(n)}\right]\left[\boldsymbol{a}^{(n)}\right]^T$$
$$- \frac{1}{N^2}\left[\sum_{n=n_0+1}^{n_0+N} \boldsymbol{a}^{(n)}\right]\left[\sum_{n=n_0+1}^{n_0+N} \boldsymbol{a}^{(n)}\right]^T \qquad (30)$$

and

$$\mathrm{Var}\{\sigma^2\,|\,\boldsymbol{Y}\} \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N}\left[\sigma^{2^{(n)}}\right]^2 - \frac{1}{N^2}\left[\sum_{n=n_0+1}^{n_0+N}\sigma^{2^{(n)}}\right]^2.$$

(31)

Note that the above computations are exact in the limit as $N \longrightarrow \infty$. However, since they involve only a finite number of samples, we think of them as approximations but realize that in theory any order of precision can be achieved given sufficiently large sample size N.

The complexity of the above Gibbs multiuser detector *per iteration* is $O(K^2 + KM)$, i.e., it has a term that is quadratic with respect to the number of users K [due to the inversion of the positive definite symmetric matrix in (20)] and a term that is *linear* with respect to the symbol block size M. The total complexity is then $O[(K^2 + KM)(n_0 + N)]$. This is a substantial complexity reduction compared with the direct implementation of the Bayesian symbol estimate (14), whose complexity is $O(2^{KM})$.

V. ADAPTIVE BAYESIAN MULTIUSER DETECTION IN IMPULSIVE NOISE

In the previous section, we assumed that the distribution of the ambient channel noise is Gaussian, as did in most of the previous work on multiuser detection. However, in many realistic communication channels, especially wireless channels, the ambient noise is known to be decidedly impulsive, due to impulsive phenomena [8], [9]. In this section, we develop the Gibbs multiuser detector in impulsive noise. It is assumed that the noise samples $\{n_j(i)\}_{j=0}^{P-1}$ of $\boldsymbol{n}(i)$ in (1) are independent with a common two-term Gaussian mixture pdf, which is given by

$$p(n_j(i)) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{n_j(i)^2}{2\sigma_1^2}\right)$$
$$+ \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{n_j(i)^2}{2\sigma_2^2}\right)$$

(32)

with $0 < \epsilon < 1$ and $\sigma_1^2 < \sigma_2^2$. This model serves as an approximation to the more fundamental Middleton Class A noise model [30], [31], [56] and has been used extensively to model physical noise arising in radar, acoustic, and mobile radio channels.

A. Prior Distributions

Define the following indicator random variable to indicate the distribution of the noise sample $n_j(i)$

$$I_j(i) = \begin{cases} 1, & \text{if } n_j(i) \sim \mathcal{N}(0, \sigma_1^2), \\ 2, & \text{if } n_j(i) \sim \mathcal{N}(0, \sigma_2^2), \\ & i = 0, \cdots, M-1; \ j = 0, \cdots, P-1. \end{cases}$$

(33)

Denote $\boldsymbol{I} \triangleq \{I_j(i)\}_{j=0;i=0}^{P-1;M-1}$, and

$$\boldsymbol{\Lambda}(i) \triangleq \mathrm{diag}\left(\sigma_{I_0(i)}^2, \sigma_{I_1(i)}^2, \cdots, \sigma_{I_{P-1}(i)}^2\right)$$
$$i = 0, \cdots, M-1.$$

(34)

The unknown quantities in this case are $(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X})$. The joint posterior distribution of these unknown quantities based on the received signal $\boldsymbol{Y}$ takes the form of

$$p(\boldsymbol{a},\,\sigma_1^2,\,\sigma_2^2,\,\epsilon,\,\boldsymbol{I},\,\boldsymbol{X}\,|\,Y)$$
$$\propto p(\boldsymbol{Y}\,|\,\boldsymbol{a},\,\sigma_1^2,\,\sigma_2^2,\,\epsilon,\,\boldsymbol{I},\,\boldsymbol{X})\,p(\boldsymbol{a})p(\sigma_1^2)\,p(\sigma_2^2)$$
$$\cdot p(\epsilon)p(\boldsymbol{I}\,|\,\epsilon)p(\boldsymbol{X})$$
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=0}^{M-1}[\boldsymbol{r}(i)-\boldsymbol{SAx}(i)]^T\boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i)-\boldsymbol{SAx}(i)]\right\}$$
$$\cdot \left(\frac{1}{\sigma_1^2}\right)^{(1/2)\sum_{i=0}^{M-1}n_1(i)}\left(\frac{1}{\sigma_2^2}\right)^{(1/2)\sum_{i=0}^{M-1}n_2(i)}$$
$$\cdot p(\boldsymbol{a})p(\sigma_1^2)\,p(\sigma_2^2)\,p(\epsilon)p(\boldsymbol{I}\,|\,\epsilon)p(\boldsymbol{X})$$

(35)

where $n_l(i)$ is the number of l's in $\{I_0(i), I_1(i), \cdots, I_{P-1}(i)\}$, $l = 1, 2$. [Note that $n_1(i) + n_2(i) = P$.] We next specify the conjugate prior distributions of the unknown quantities in (35).

As in the case of Gaussian noise, the prior distributions $p(\boldsymbol{a})$ and $p(\boldsymbol{X})$ are given, respectively, by (15) and (18). For the noise variances $\sigma_l^2, l = 1, 2$, independent inverse chi-square distributions are assumed, i.e.,

$$p(\sigma_l^2) \sim \chi^{-2}(\nu_l, \lambda_l), \qquad l = 1, 2, \text{ with } \nu_1\lambda_1 < \nu_2\lambda_2.$$ (36)

For the impulse probability ϵ, a prior of Beta distribution is assumed, i.e.,

$$p(\epsilon) = \frac{\Gamma(a_0+b_0)}{\Gamma(a_0)\Gamma(b_0)}\,\epsilon^{a_0-1}(1-\epsilon)^{b_0-1} \sim \mathrm{Beta}(a_0, b_0).$$ (37)

Note that the value $a_0/(a_0+b_0)$ reflects the prior knowledge of the value of ϵ. Moreover, $(a_0 + b_0)$ reflects the strength of the prior belief, i.e., roughly the prior knowledge is worth (a_0+b_0) data points. Given ϵ, the conditional distribution of the indicator $I_j(i)$ is then

$$P[I_j(i) = 1\,|\,\epsilon] = 1 - \epsilon, \quad\text{and}\quad P[I_j(i) = 2\,|\,\epsilon] = \epsilon$$ (38)
$$\Rightarrow p(\boldsymbol{I}\,|\,\epsilon) = (1-\epsilon)^{m_1}\epsilon^{m_2}$$ (39)

with

$$m_1 \triangleq \sum_{i=0}^{M-1}n_1(i)$$

$$m_2 \triangleq \sum_{i=0}^{M-1}n_2(i) = MP - m_1.$$

B. Conditional Posterior Distributions

The following conditional posterior distributions are required by the Gibbs multiuser detector in impulsive noise. The derivations are found in Appendix B.

1) The conditional distribution of the amplitude vector $\boldsymbol{a}$ given $\sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}$, and $\boldsymbol{Y}$ is

$$p(\boldsymbol{a}\,|\,\sigma_1^2,\,\sigma_2^2,\,\epsilon,\,\boldsymbol{I},\,\boldsymbol{X},\,Y) \sim \mathcal{N}(\boldsymbol{a}_*, \boldsymbol{\Sigma}_*)I_{\{\boldsymbol{a}>\boldsymbol{0}\}}$$ (40)

with

$$\boldsymbol{\Sigma}_*^{-1} \triangleq \boldsymbol{\Sigma}_0^{-1} + \sum_{i=0}^{M-1}\boldsymbol{B}(i)\boldsymbol{S}^T\boldsymbol{\Lambda}(i)^{-1}\boldsymbol{S}\boldsymbol{B}(i)$$ (41)

and

$$\boldsymbol{a}_* \triangleq \boldsymbol{\Sigma}_*\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{a}_0 + \sum_{i=0}^{M-1}\boldsymbol{B}(i)\boldsymbol{S}^T\boldsymbol{\Lambda}(i)^{-1}\boldsymbol{r}(i)\right).$$ (42)

2) The conditional distribution of the noise variance σ_l^2 given $\boldsymbol{a}$, $\sigma_{\bar{l}}^2$, ϵ, $\boldsymbol{I}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$ is given by [here $\bar{l} = 2$ if $l = 1$, and $\bar{l} = 1$ if $l = 2$].

$$p\left(\sigma_l^2 \,|\, \boldsymbol{a}, \sigma_{\bar{l}}^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)$$
$$\sim \chi^{-2}\left(\left[\nu_l + \sum_{i=0}^{M-1} n_l(i)\right], \frac{\nu_l \lambda_l + s_l^2}{\nu_l + \sum_{i=0}^{M-1} n_l(i)}\right) \quad (43)$$

with

$$s_l^2 \triangleq \sum_{i=0}^{M-1} \sum_{j=0}^{P-1}\left[r_j(i) - \boldsymbol{\xi}_j^T \boldsymbol{A} \boldsymbol{x}(i)\right]^2 \cdot \mathbf{1}_{\{I_j(i)=l\}}$$
$$l = 1, 2. \quad (44)$$

In (44), $\mathbf{1}_{\{I_j(i)=l\}}$ is the indicator function such that it is 1 if $I_j(i) = l$, and it is 0 if $I_j(i) \neq l$; $\boldsymbol{\xi}_j^T$ is the jth row of the spreading waveform matrix $\boldsymbol{S}$, $j = 0, \cdots, P-1$.

3) The conditional probability of $x_k(i) = \pm 1$, given $\boldsymbol{a}$, σ_1^2, σ_2^2, ϵ, $\boldsymbol{I}$, $\boldsymbol{X}_{ki}$, and $\boldsymbol{Y}$ can be obtained from [where $\boldsymbol{X}_{ki}$ denotes the set containing all elements of $\boldsymbol{X}$ except for $x_k(i)$]

$$\frac{P[x_k(i) = +1 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{ki}, \boldsymbol{Y}]}{P[x_k(i) = -1 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{ki}, \boldsymbol{Y}]}$$
$$= \frac{\rho_k(i)}{1 - \rho_k(i)} \cdot \exp\left\{2A_k \boldsymbol{s}_k^T \boldsymbol{\Lambda}(i)^{-1}\left[\boldsymbol{r}(i) - \boldsymbol{S} \boldsymbol{A} \boldsymbol{x}^0(i)\right]\right\}$$
$$k = 1, \cdots, K; \; i = 0, \cdots, M-1 \quad (45)$$

where
$$\boldsymbol{x}_k^0(i) \triangleq [x_1(i), \cdots, x_{k-1}(i), 0, x_{k+1}(i), \cdots, x_K(i)]^T.$$

4) The conditional distribution of $I_j(i)$, given $\boldsymbol{a}$, σ_1^2, σ_2^2, ϵ, $\boldsymbol{I}_{ji}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$ is given by [$\boldsymbol{I}_{ji}$ denotes the set containing all elements of $\boldsymbol{I}$ except for $I_j(i)$]

$$\frac{P[I_j(i) = 1 | \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{ji}, \boldsymbol{X}, \boldsymbol{Y}]}{P[I_j(i) = 2 | \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{ji}, \boldsymbol{X}, \boldsymbol{Y}]}$$
$$= \frac{1 - \epsilon}{\epsilon}\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{1/2} \exp\left\{\frac{1}{2}\left[r_j(i) - \boldsymbol{\xi}_j^T \boldsymbol{A} \boldsymbol{x}(i)\right]^2\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right\}$$
$$j = 0, \cdots, P-1; \; i = 0, \cdots, M-1. \quad (46)$$

5) The conditional distribution of ϵ, given $\boldsymbol{a}$, σ_1^2, σ_2^2, $\boldsymbol{I}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$ is given by

$$p\left(\epsilon \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)$$
$$= \text{Beta}\left(a_0 + \sum_{i=0}^{M-1} n_2(i), b_0 + \sum_{i=0}^{M-1} n_1(i)\right). \quad (47)$$

C. Gibbs Multiuser Detector in Impulsive Noise

Using the above conditional posterior distributions, the Gibbs sampling implementation of the adaptive multiuser detector in impulsive noise proceeds iteratively as follows. Given initial values of the unknown quantities $\{\boldsymbol{a}^{(0)}, \sigma_1^{2^{(0)}}, \sigma_2^{2^{(0)}}, \epsilon^{(0)}, \boldsymbol{I}^{(0)}, \boldsymbol{X}^{(0)}\}$ drawn from their prior distributions, and for $n = 1, 2, \cdots$, we have the following.

1) Draw $\boldsymbol{a}^{(n)}$ from $p(\boldsymbol{a} \,|\, \sigma_1^{2^{(n-1)}}, \sigma_2^{2^{(n-1)}}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (40).

2) Draw $\sigma_1^{2^{(n)}}$ from $p(\sigma_1^2 \,|\, \boldsymbol{a}^{(n)}, \sigma_2^{2^{(n-1)}}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (43);
Draw $\sigma_2^{2^{(n)}}$ from $p(\sigma_2^2 \,|\, \boldsymbol{a}^{(n)}, \sigma_1^{2^{(n)}}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (43).

3) For $i = 0, 1, \cdots, M-1$
For $k = 1, 2, \cdots, K$
Draw $x_k(i)^{(n)}$ from $P[x_k(i) \,|\, \boldsymbol{a}^{(n)}, \sigma_1^{2^{(n)}}, \sigma_2^{2^{(n)}}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}_{ki}^{(n)}, \boldsymbol{Y}]$ given by (45), where

$$\boldsymbol{X}_{ki}^{(n)} \triangleq \left\{\boldsymbol{x}(0)^{(n)}, \cdots, \boldsymbol{x}(i-1)^{(n)}, x_1(i)^{(n)}, \cdots, \right.$$
$$x_{k-1}(i)^{(n)}, x_{k+1}(i)^{(n-1)}, \cdots, x_K(i)^{(n-1)},$$
$$\left.\boldsymbol{x}(i+1)^{(n-1)}, \cdots, \boldsymbol{x}(M-1)^{(n-1)}\right\}.$$

4) For $i = 0, 1, \cdots, M-1$
For $j = 1, 2, \cdots, P-1$
Draw $I_j(i)^{(n)}$ from $P[I_j(i) \,|\, \boldsymbol{a}^{(n)}, \sigma_1^{2^{(n)}}, \sigma_2^{2^{(n)}}, \epsilon^{(n-1)}, \boldsymbol{I}_{ji}^{(n)}, \boldsymbol{X}^{(n)}, \boldsymbol{Y}]$ given by (46), where

$$\boldsymbol{I}_{ji}^{(n)} \triangleq \left\{I_0(0)^{(n)}, \cdots, I_{P-1}(0)^{(n)}, \cdots, I_0(i-1)^{(n)}, \cdots\right.$$
$$I_{P-1}(i-1)^{(n)}, I_0(i)^{(n)}, \cdots, I_{j-1}(i)^{(n)},$$
$$I_{j+1}(i)^{(n-1)}, \cdots, I_{P-1}(i)^{(n-1)}, \cdots,$$
$$\left.I_{P-1}(M-1)^{(n-1)}\right\}.$$

5) Draw $\epsilon^{(n)}$ from $p(\epsilon \,|\, \boldsymbol{a}^{(n)}, \sigma_1^{2^{(n)}}, \sigma_2^{2^{(n)}}, \boldsymbol{I}^{(n)}, \boldsymbol{X}^{(n)}, \boldsymbol{Y})$ given by (47).

As in the case of Gaussian noise, the *a posteriori* symbol probabilities $P[x_k(i) = +1 \,|\, \boldsymbol{Y}]$ are computed using (26). The *a posteriori* means and variances of the other unknown quantities can also be computed similar to (28)–(31).

The complexity of the above Gibbs multiuser detector is $O[(K^2 + KM + PM)]$ per iteration. Note that the direct implementation of the Bayesian symbol estimate based on (35) has a computational complexity of $O(2^{KM+PM})$.

VI. ITERATIVE JOINT MULTIUSER DETECTION AND DECODING—ADAPTIVE TURBO MULTIUSER DETECTION

Since the discovery of the powerful Turbo codes [6], [7], considerable attention has been redrawn to iterative ("Turbo") processing techniques. The so-called Turbo principle can be successfully applied to many detection/decoding problems such as serial concatenated decoding, equalization, coded modulation, multiuser detection, and joint source and channel decoding [21]. In this section, we consider employing iterative joint multiuser detection and decoding to improve the performance of the adaptive Bayesian multiuser detector in a coded CDMA system. Because it utilizes the *a priori* symbol probabilities, and it produces symbol (or bit) *a posteriori* probabilities, the adaptive Bayesian multiuser detectors developed in this paper is well suited for iterative processing, which allows the adaptive multiuser detector to refine its processing based on the information from the decoding stage, and vice versa. In [51], a Turbo multiuser receiver is developed for coded CDMA systems with Gaussian noise, under the assumption that the received amplitudes and the noise variance are *known* to the receiver. In what follows, we develop
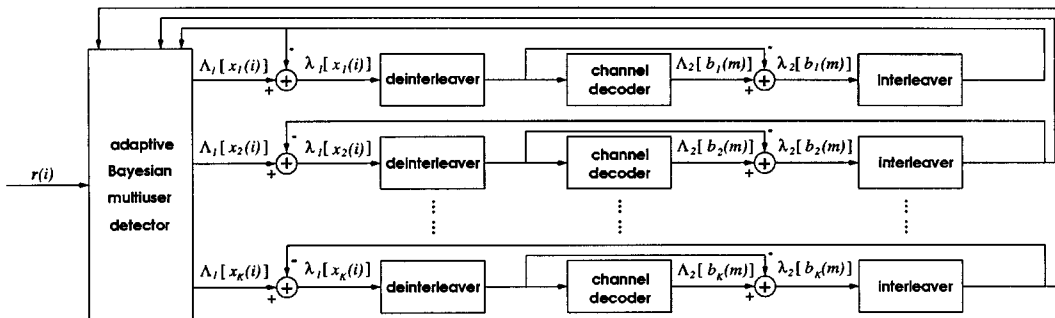
Fig. 2. Iterative processing for joint Bayesian multiuser detection and decoding—adaptive Turbo multiuser detection.

adaptive Turbo multiuser receivers for both Gaussian and impulsive noise CDMA channels, with *unknown* amplitudes and noise parameters.

The iterative (Turbo) receiver structure is shown in Fig. 2. It consists of two stages: i) the adaptive multiuser detector developed in the previous sections followed by ii) a soft-input soft-output channel decoder. The two stages are separated by deinterleavers and interleavers. As discussed in the previous sections, the adaptive multiuser detector delivers the *a posteriori* symbol probabilities $\{P[x_k(i) = +1 \mid Y]\}_{k=1; i=0}^{K; M-1}$. Based on these, we first compute the *a posteriori* log-likelihood ratios of a transmitted "+1" symbol and a transmitted "−1" symbol

$$\Lambda_1[x_k(i)] \triangleq \log \frac{P[x_k(i) = +1 \mid Y]}{P[x_k(i) = -1 \mid Y]}, \qquad k = 1, \cdots, K$$
$$i = 0, \cdots, M-1. \tag{48}$$

Using the Bayes' rule, (48) can be written as

$$\Lambda_1[x_k(i)] = \underbrace{\log \frac{p[Y \mid x_k(i) = +1]}{p[Y \mid x_k(i) = -1]}}_{\lambda_1[x_k(i)]} + \underbrace{\log \frac{P[x_k(i) = +1]}{P[x_k(i) = -1]}}_{\lambda_2^p[x_k(i)]} \tag{49}$$

where the second term in (49), which is denoted by $\lambda_2^p[x_k(i)]$, represents the *a priori* LLR of the code bit $x_k(i)$, which is computed by the channel decoder in the previous iteration, interleaved, and then fed back to the adaptive Bayesian multiuser detector. (The superscript p indicates the quantity obtained from the previous iteration). For the first iteration, assuming equally likely code bits, i.e., no prior information available, we then have $\lambda_2^p[x_k(i)] = 0$, $k = 1, \cdots, K$, $i = 0, \cdots, M-1$. The first term in (49), which is denoted by $\lambda_1[x_k(i)]$, represents the *extrinsic* information delivered by the adaptive Bayesian multiuser detector, based on the received signals Y, the structure of the multiuser signal given by (1), and the prior information about all other code bits. The extrinsic information $\lambda_1[x_k(i)]$, which is not influenced by the *a priori* information $\lambda_2^p[x_k(i)]$ provided by the channel decoder, is then reverse interleaved and fed into the channel decoder, as the *a priori* information in the next iteration.

Based on the extrinsic information of the code bits $\{\lambda_1^p[x_k(m)]\}_{k=1; m=0}^{K; M-1}$ and the structure of the channel code, the soft-input soft-output channel decoder computes the *a posteriori* LLR of each code bit [51], as shown in (50), shown at the bottom of the page. It is seen from (50) that the output of the soft-input soft-output channel decoder is the sum of the prior information $\lambda_1^p[x_k(m)]$, and the *extrinsic* information $\lambda_2[x_k(m)]$ delivered by the channel decoder. This extrinsic information is the information about the code bit $x_k(m)$ gleaned from the prior information about the other code bits $\{\lambda_1^p[x_k(l)]\}_{l \neq m}$ based on the constraint structure of the code. The soft channel decoder also computes the *a posteriori* LLR of every information bit, which is used to make decision on the decoded bit at the last iteration. After interleaving, the extrinsic information delivered by the channel decoder $\{\lambda_2[x_k(m)]\}_{k=1; m=0}^{K; M-1}$ is then used to compute the *a priori* symbol distributions $\{\rho_k(i)\}_{k=1; i=0}^{K; M-1}$ defined in (7) from the corresponding LLR's as follows. Since

$$\lambda_2^p[x_k(i)] = \log \frac{P[x_k(i) = +1]}{P[x_k(i) = -1]}$$

after some manipulations, we have

$$\begin{aligned}
\rho_k(i) &\triangleq P[x_k(i) = +1] = \frac{\exp(\lambda_2^p[x_k(i)])}{1 + \exp(\lambda_2^p[x_k(i)])} \\
&= \frac{\exp\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)}{\exp\left(-\frac{1}{2}\lambda_2^p[x_k(i)]\right) + \exp\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)} \\
&= \frac{\cosh\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)\left[1 + \tanh\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)\right]}{2\cosh\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)} \\
&= \frac{1}{2}\left[1 + \tanh\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)\right]. \tag{51}
\end{aligned}$$

The symbol probabilities $\{\rho_k(i)\}_{k=1; i=0}^{K; M-1}$ are then fed back to the adaptive Bayesian multiuser detector as the prior information for the next iteration. Note that at the first iteration, the extrinsic information $\{\lambda_1[x_k(i)]\}$ and $\{\lambda_2[x_k(i)]\}$ are statistically independent. However, subsequently, since they use the same information indirectly, they will become increasingly correlated, and finally, the improvement through the iterations will diminish.

$$\Lambda_2[x_k(m)] \triangleq \log \frac{P[x_k(m) = +1 \mid \{\lambda_1^p[x_k(i)]\}_{k=1; i=0}^{K; M-1}; \text{ code constraints}]}{P[x_k(m) = -1 \mid \{\lambda_1^p[x_k(i)]\}_{k=1; i=0}^{K; M-1}; \text{ code constraints}]} = \lambda_2[x_k(m)] + \lambda_1^p[x_k(m)] \tag{50}$$

VII. Discussions

A. Decoder-Assisted Convergence Assessment

Detecting convergence in the Gibbs sampler is usually done in some *ad hoc* way. Some methods can be found in [44]. One of them is to monitor a sequence of weights that measure the discrepancy between the sampled and the desired distribution. In the application considered here, since the adaptive multiuser detector is followed by a bank of channel decoders, we can assess convergence by monitoring the number of bit corrections made by the channel decoders. If this number exceeds some predetermined threshold, then we decide that convergence is not achieved. In that case, the Gibbs multiuser detector will be applied again to the same data block. The rationale is that if the Gibbs sampler has reached convergence, then the symbol (and bit) errors after multiuser detection should be relatively small. On the other hand, if convergence is not reached, then the code bits generated by the multiuser detector are virtually random and do not satisfy the constraints imposed by the code trellises. Hence, the channel decoders will make a large number of corrections. Note that there is no additional computational complexity for such a convergence detection. We only need to compare the signs of the code-bit log-likelihood ratios at the input and the output of the soft channel decoder to determine the number of corrections made.

B. Code-Constrained Gibbs Multiuser Detectors

Another approach to exploiting the coded signal structure in adaptive Bayesian multiuser detection is to make use of the code constraints in the Gibbs sampler. For instance, suppose that the user information bits are encoded by some *block code* of length L and that the code bits are *not* interleaved. Then, one signal frame of M symbols contains $J = M/L$ code words, with the jth code word given by

$$\underline{x}_k(j) = [x_k(jL), x_k(jL+1), \cdots, x_k(jL+L-1)]$$
$$j = 0, 1, \cdots, \frac{M}{L} - 1, \ k = 1, \cdots, K.$$

Let $\mathcal{X}_k$ be the set of all valid code words for user k. Now, in the Gibbs sampler, instead of drawing each individual symbol $x_k(i)$ once a time according to (25) or (45), we draw a code word $\underline{x}_k(j)$ of L symbols from $\mathcal{X}_k$ each time. Specifically, let $-\underline{1}$ denote the code word with all entries being "-1"s (this is the so-called all-zero code word, and it is always a valid code word for any block code [53]). If the ambient channel noise is Gaussian, then for any code word $\underline{u} \in \mathcal{X}_k$, the conditional probability of $\underline{x}_k(j) = \underline{u}$, given the values of the rest of the unknowns, can be obtained from

$$\frac{P[\underline{x}_k(j) = \underline{u} \,|\, \boldsymbol{a}, \sigma^2, \boldsymbol{X}_{kj}, \boldsymbol{Y}]}{P[\underline{x}_k(j) = -\underline{1} \,|\, \boldsymbol{a}, \sigma^2, \boldsymbol{X}_{kj}, \boldsymbol{Y}]}$$
$$= \frac{\rho_{kj}(\underline{u})}{1 - \rho_{kj}(\underline{u})}$$
$$\cdot \exp\left\{ \frac{2A_k}{\sigma^2} \boldsymbol{s}_k^T \sum_{\substack{l=0 \\ u(l) \neq -1}}^{L-1} \left[\boldsymbol{r}(jL+l) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}_k^0(jL+l) \right] \right\}$$
$$k = 1, \cdots, K; \ j = 0, 1, \cdots, \frac{M}{L} - 1 \qquad (52)$$

where $\boldsymbol{X}_{kj}$ denotes the set containing all elements of $\boldsymbol{X}$, except for $\underline{x}_k(j)$; $\rho_{kj}(\underline{u}) \triangleq P[\underline{x}_k(j) = \underline{u}]$; and $\boldsymbol{x}_k^0(i) \triangleq [x_1(i), \cdots, x_{k-1}(i), 0, x_{k+1}(i), \cdots, x_K(i)]^T$. On the other hand, if the ambient channel noise is impulsive, we have

$$\frac{P[\underline{x}_k(j) = \underline{u} \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{kj}, \boldsymbol{Y}]}{P[\underline{x}_k(j) = -\underline{1} \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{kj}, \boldsymbol{Y}]}$$
$$= \frac{\rho_{kj}(\underline{u})}{1 - \rho_{kj}(\underline{u})} \cdot \exp\left\{ 2A_k \boldsymbol{s}_k^T \sum_{\substack{l=0 \\ u(l) \neq -1}}^{L-1} \boldsymbol{\Lambda}(jL+l)^{-1} \right.$$
$$\left. \cdot \left[\boldsymbol{r}(jL+l) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}^0(jL+l) \right] \right\}$$
$$k = 1, \cdots, K; \ j = 0, \cdots, \frac{M}{L} - 1. \qquad (53)$$

The conditional distributions for sampling the other unknowns remain the same as before. The advantage of sampling a code word instead of sampling an individual symbol is that it can significantly improve the accuracy of samples drawn by the Gibbs sampler since only valid code words are drawn. This will be demonstrated by some simulation examples in the next section.

C. Relationship Between the Gibbs Sampler and the EM Algorithm

The expectation-maximization (EM) algorithm has also been applied to joint parameter estimation and multiuser detection [36]. The major advantage of the Gibbs sampling technique over the EM algorithm is that the Gibbs sampler is a *global* optimization technique. The EM algorithm is a *local* optimization method, and it can easily get trapped by local extrema on the likelihood surface. The EM method performs well if the initial estimates of the channel and symbols are close to their true values. On the other hand, the Gibbs sampler is guaranteed to converge to the global optimum with any random initialization. Of course, the convergence rate crucially depends on the shape the joint posterior density surface. When the posterior distribution has several modes separated by very low density regions (energy gap), then the Gibbs sampler that generates "random walks" according to the distribution may have difficulties crossing such gaps in visiting all modes. If such a gap is severe, then the random walk may get stuck within one mode for a long time before it moves to another mode. Many modifications of the Gibbs sampler have been developed to combat the "large energy gap" situation. For example, see [20] and [54].

VIII. Simulations

A. Simulation Setup

In this section, we provide a number of simulation examples to illustrate the performance of the adaptive Bayesian multiuser detectors developed in this paper. We consider a five-user ($K = 5$) synchronous CDMA channel with processing gain $P = 10$.

The user-spreading waveform matrix S and the corresponding correlation matrix R are given, respectively, by

$$S^T = \frac{1}{\sqrt{10}} \begin{bmatrix} -1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 \\ -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix},$$

$$R = S^T S = \frac{1}{10} \begin{bmatrix} 10 & -2 & -2 & 4 & -2 \\ -2 & 10 & 2 & 0 & 2 \\ -2 & 2 & 10 & -4 & 2 \\ 4 & 0 & -4 & 10 & -4 \\ -2 & 2 & 2 & -4 & 10 \end{bmatrix}.$$

In all the simulations described in this section, the following *noninformative conjugate* prior distributions are used in the Gibbs sampler. For the case of Gaussian noise

$$p\left(\boldsymbol{a}^{(0)}\right) \sim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{\Sigma}_0) I_{\{\boldsymbol{a}^{(0)} > 0\}} \longrightarrow \boldsymbol{a}_0 = [1\ 1\ 1\ 1\ 1]^T$$
$$\boldsymbol{\Sigma}_0 = 1000\boldsymbol{I}$$
$$p\left(\sigma^{2(0)}\right) \sim \chi^{-2}(\nu_0, \lambda_0) \longrightarrow \nu_0 = 1, \qquad \lambda_0 = 0.1$$

and for the case of impulsive noise

$$p\left(\boldsymbol{a}^{(0)}\right) \sim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{\Sigma}_0) I_{\{\boldsymbol{a}^{(0)} > \boldsymbol{0}\}} \longrightarrow \boldsymbol{a}_0 = [1\ 1\ 1\ 1\ 1]^T$$
$$\boldsymbol{\Sigma}_0 = 1000\boldsymbol{I}$$
$$p\left(\sigma_1^{2(0)}\right) \sim \chi^{-2}(\nu_1, \lambda_1) \longrightarrow \nu_1 = 1, \qquad \lambda_1 = 0.1$$
$$p\left(\sigma_2^{2(0)}\right) \sim \chi^{-2}(\nu_1, \lambda_1) \longrightarrow \nu_1 = 1, \qquad \lambda_1 = 1$$
$$p\left(\epsilon^{(0)}\right) \sim \text{Beta}(a_0, b_0) \longrightarrow a_0 = 1, \qquad b_0 = 2.$$

Note that the performance of the Gibbs sampler is insensitive to the values of the parameters in these priors, as long as the priors are noninformative.

In adaptive Turbo multiuser detection, for the first iteration, the prior symbol probabilities $\rho_k(i) \triangleq P[x_k(i) = +1] = 1/2$ for all symbols; in the subsequent iterations, the prior symbol probabilities are provided by the channel decoder, as given by (51). The data block size of each user is $M = 256$. For each data block, the Gibbs sampling is performed for 100 iterations, with the first 50 iterations as the "burn-in" period. The decoder-assisted convergence assessment is employed. Specifically, if the number of bit corrections made by the decoder exceeds one third of the total number of bits (i.e., $M/3$), then it is decided that convergence is not reached, and the Gibbs sampler is applied to the same data block again.

B. Convergence Behavior of the Gibbs Multiuser Detectors

Example 1: We first illustrate the convergence behavior of the proposed adaptive Bayesian multiuser detector in Gaussian noise. In this example, the user amplitudes and the noise variance are given by

$$A_1^2 = -4 \text{ dB}, \qquad A_2^2 = -2 \text{ dB}, \qquad A_3^3 = 0 \text{ dB}$$
$$A_4^2 = 2 \text{ dB}, \qquad A_5^2 = 4 \text{ dB}, \qquad \sigma^2 = -2 \text{ dB}.$$

In Fig. 3, we plot the first 100 samples drawn by the Gibbs sampler of the parameters $x_3(50), x_4(100), A_1, A_5$, and σ^2. The corresponding true values of these quantities are also shown in the same figure as the straight lines. Note that in this case, the
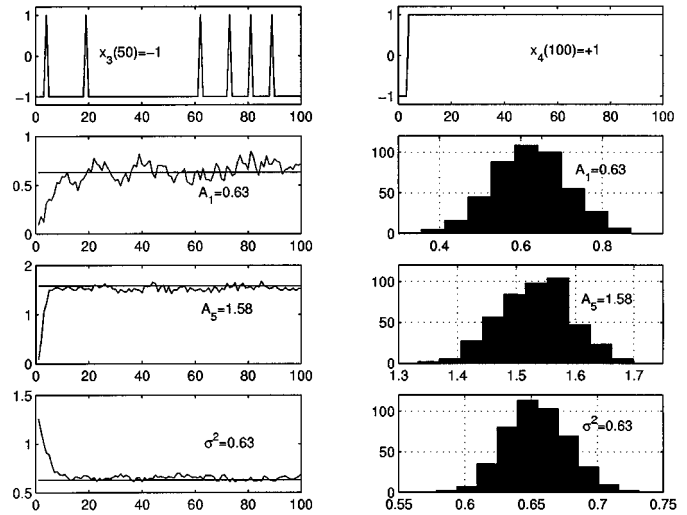

Fig. 3.   Samples and histograms—Gaussian noise. $A_1^2 = -4$ dB, $A_2^2 = -2$ dB, $A_3^2 = 0$ dB, $A_4^2 = 2$ dB, $A_5^2 = 4$ dB, and $\sigma^2 = -2$ dB. The histograms are based on 500 samples collected after the initial 50 iterations.
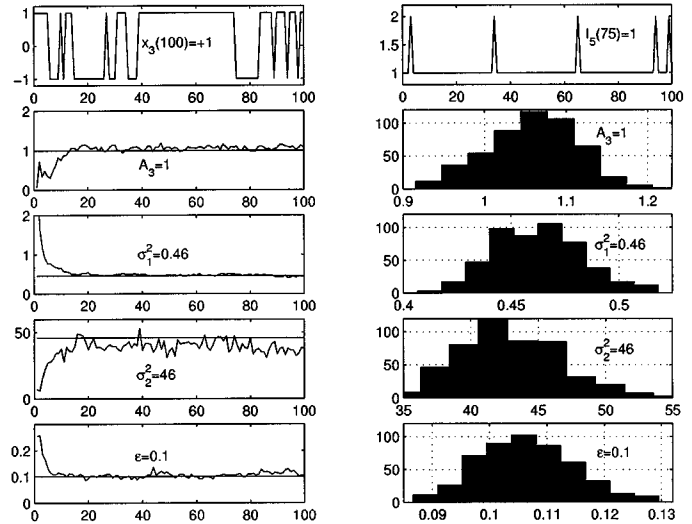

Fig. 4.   Samples and histograms—impulsive noise. $A_1^2 = -4$ dB, $A_2^2 = -2$ dB, $A_3^2 = 0$ dB, $A_4^2 = 2$ dB, $A_5^2 = 4$ dB, $\epsilon = 0.1$, $\sigma_2^2/\sigma_1^2 = 100$, $\sigma^2 \triangleq (1 - \epsilon)\sigma_1^2 + \epsilon\sigma_2^2 = 7$ dB. The histograms are based on 500 samples collected after the initial 50 iterations.

number of unknown parameters is $(K + KM + 1)$ (i.e., $\boldsymbol{a}, \boldsymbol{X}$, and σ^2). Remarkably, it is seen that the Gibbs sampler reaches convergence within about 20 iterations. The marginal posterior distributions of the unknown parameters A_1, A_5, and σ^2 in the steady state can be illustrated by the corresponding histograms, which are also shown in Fig. 3. The histograms are based on 500 samples collected after the initial 50 iterations.

Example 2: We next illustrate the convergence behavior of proposed adaptive Bayesian multiuser detector in impulsive noise. The user amplitudes are the same as in Example 1. The noise samples are generated according to the two-term Gaussian model (5) with the following parameters:

$$\epsilon = 0.1, \quad \sigma_2^2/\sigma_1^2 = 100, \quad \sigma^2 \triangleq (1 - \epsilon)\sigma_1^2 + \epsilon\sigma_2^2 = 7 \text{ dB}.$$

The first 100 samples drawn by the Gibbs sampler of the parameters $x_3(100), I_5(75), A_3, \sigma_1^2, \sigma_2^2$, and ϵ are shown in Fig. 4. The corresponding true values of these quantities are also shown in
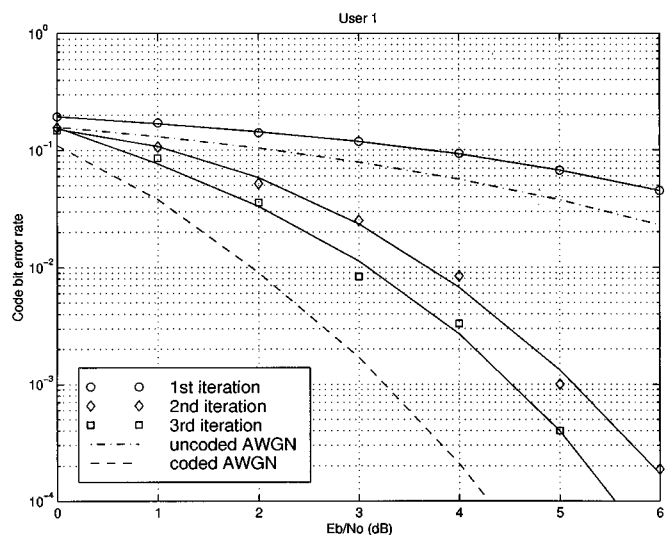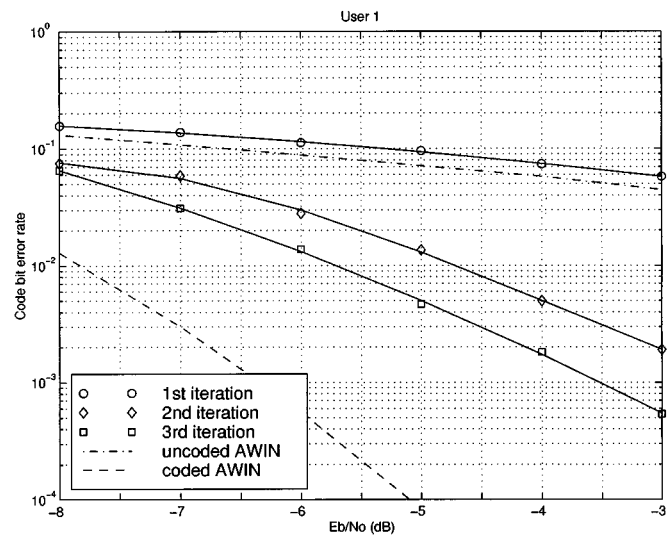
(a)
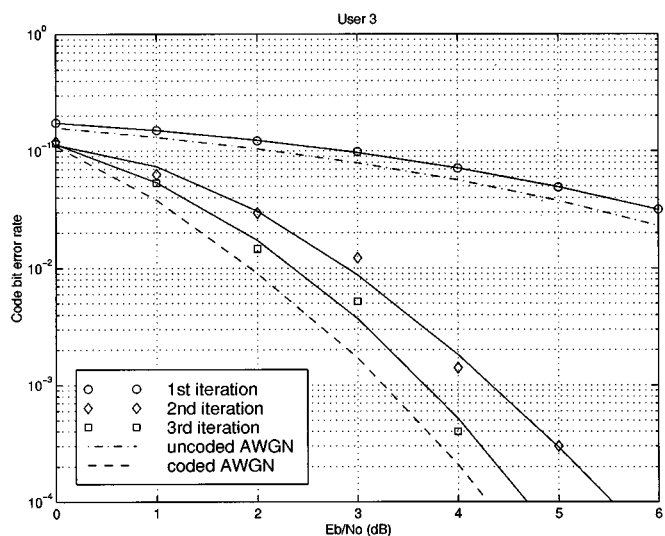
(b)

Fig. 5. Bit error rate performance—convolutional code, Gaussian noise. All users have the same amplitudes.
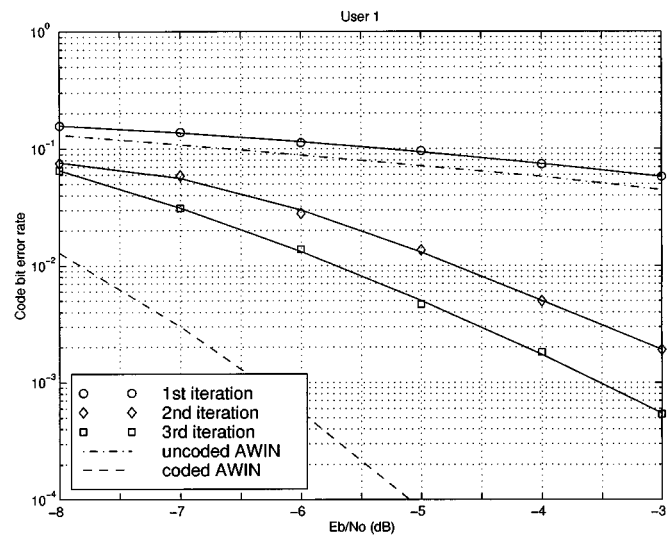

(a)

(b)

Fig. 6. Bit error rate performance—convolutional code, impulsive noise. All users have the same amplitudes. $\sigma_2^2/\sigma_1^2 = 100$ and $\epsilon = 0.1$.

the same figure as the straight lines. Note that in this case, the number of unknown parameters is $(K+KM+PM+3)$ (i.e., $\boldsymbol{a}$, $\boldsymbol{X}, \boldsymbol{I}, \sigma_1^2, \sigma_2^2$ and ϵ). It is seen that as in the Gaussian noise case, the Gibbs sampler converges within about 20 samples. The histograms of the unknown parameters $A_3, \sigma_1^2, \sigma_2^2$, and ϵ are also shown in Fig. 4, which are based on 500 samples collected after the initial 50 iterations.

C. Performance of the Adaptive Turbo Multiuser Detectors

Examples 3 and 4: We now illustrate the performance of the adaptive Turbo multiuser detectors discussed in Section VI. The channel code for each user is a rate that is one half of the constraint length-5 convolutional code (with generators 23, 35 in octal notation). The interleaver of each user is independently and randomly generated and fixed for all simulations. The block size of the information bits is 128. (i.e., the code bit block size is $M = 256$.) The code bits are BPSK modulated, i.e., $x_k \in \{+1, -1\}$. All users have the same signal-to-noise

ratio (E_b/N_0). The symbol posterior probabilities are computed according to (26) with $n_0 = N = 50$.

Fig. 5 illustrates the bit error rate performance of the adaptive Turbo multiuser detector for User 1 and User 3. The code bit error rate at the output of the adaptive Bayesian multiuser detector is plotted for the first three iterations. The curve corresponding to the first iteration is the uncoded bit error rate at the output of the adaptive Bayesian multiuser detector. The uncoded and coded bit error rate curves in a single-user additive white Gaussian noise (AWGN) channel are also shown in the same figure (as, respectively, are the dash-dotted and the dashed lines). It is seen that by incorporating the extrinsic information provided by the channel decoder as the prior symbol probabilities, the proposed adaptive Turbo multiuser detector approaches the single-user performance in an AWGN channel within a few iterations. The bit error rate performance of the adaptive Turbo multiuser detector in impulsive noise is illustrated in Fig. 6, where the code bit error rates at the output of the adaptive Bayesian multiuser detector for the first three itera-
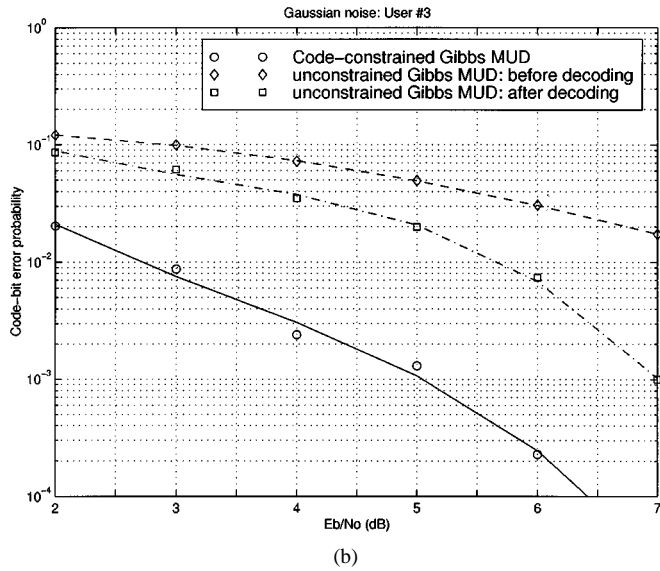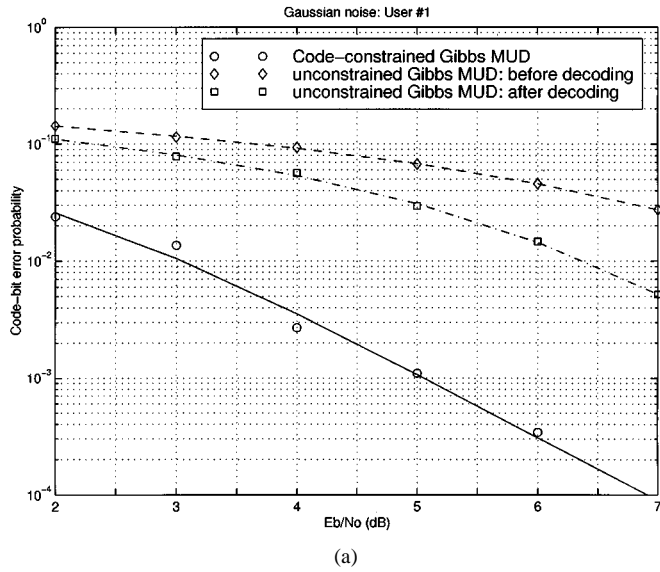
Fig. 7. Bit error rate performance—block code, Gaussian noise. All users have the same amplitudes.

Fig. 8. Bit error rate performance—block code, impulsive noise. All users have the same amplitudes. $\sigma_2^2/\sigma_1^2 = 100$ and $\epsilon = 0.1$.

tions are shown. The uncoded and coded bit error rate curves in a single-user additive white impulsive noise (AWIN) channel are also shown in the same figure (as the dash-dotted and dashed lines, respectively), where the conventional matched-filter receiver is employed for demodulation. Note that at high E_b/N_0, the performance of user 3 after the first iteration is actually better than the single-user performance. This is because the matched-filter receiver is not the optimal single-user receiver in impulsive noise. Indeed, when $K = 1$, the maximum likelihood detector for signal model (1) is given by

$$\hat{x}_1(i) = \text{sign}\left\{\sum_{j=1}^{P} \frac{r_j(i)s_{k,j}}{\sigma_j^2(i)}\right\}.$$

D. Performance of the Code-Constrained Gibbs Multiuser Detectors

Examples 5 and 6: Finally, we consider the performance of the code-constrained Gibbs multiuser detectors discussed in

Section VII. We assume that each user employs the $(7, 4)$ cyclic block code with eight possible codewords [53]:

$$\mathcal{X} = \left\{ \begin{array}{ccccccc} (& -1 & -1 & -1 & -1 & -1 & -1 & -1 &) \\ (& 1 & -1 & 1 & 1 & 1 & -1 & -1 &) \\ (& 1 & 1 & 1 & -1 & -1 & 1 & -1 &) \\ (& -1 & 1 & 1 & 1 & -1 & -1 & 1 &) \\ (& 1 & 1 & -1 & -1 & 1 & -1 & 1 &) \\ (& -1 & -1 & 1 & -1 & 1 & 1 & 1 &) \\ (& -1 & 1 & -1 & 1 & 1 & 1 & -1 &) \\ (& 1 & -1 & -1 & 1 & -1 & 1 & 1 &) \end{array} \right\}.$$

The bit error rate performance of the code-constrained Gibbs multiuser detector in Gaussian noise is shown in Fig. 7. In this case, the Gibbs sampler draws a codeword from $\mathcal{X}$ at each time, according to (52). In the same figure, the unconstrained Gibbs multiuser detector performance before and after decoding is also plotted. It is seen that by exploiting the code constraints in the Gibbs sampler, significant performance gain is achieved. The performance of the code-constrained Gibbs multiuser detector

in impulsive noise is shown in Fig. 8, and similar performance gain over the unconstrained Gibbs multiuser detector is evident.

IX. CONCLUSIONS

In this paper, we have developed a new adaptive multiuser detection scheme that is optimal in the sense that it is based on the *Bayesian* inference of all unknown quantities. Such an adaptive Bayesian multiuser detector can be efficiently implemented using the Gibbs sampler, which is a Markov chain Monte Carlo procedure for computing Bayesian estimates. We have derived the adaptive multiuser detection algorithms for both the Gaussian noise and the impulsive noise synchronous CDMA channel. The proposed adaptive Bayesian multiuser detectors can incorporate the *a priori* symbol probabilities, and they produce as output the *a posteriori* symbol probabilities. That is, they are "soft-in soft-output" algorithms. Hence, they are very well suited for iterative processing in a coded system, which allows the adaptive Bayesian multiuser detector to refine its processing based on the information from the decoding stage, and vice versa—a receiver structure termed as *adaptive Turbo multiuser detector*. Furthermore, the channel decoder facilitates a simple way of assessing the convergence of the adaptive multiuser detector by monitoring the number of bit corrections made. Moreover, if the user data are encoded by a short block code, then by exploiting the constraint on the valid code words in the Gibbs sampler, significant improvement on the performance of the adaptive Bayesian multiuser detector can be obtained. Finally, we notice that the CDMA signal model (1) treated here is a rather simple one, namely, it is real-valued and synchronous. Future extensions to this work include generalizations of the techniques proposed here to asynchronous CDMA systems and to systems with multipath fading effects.

APPENDIX A

Derivation of (19):

$$p(\boldsymbol{a}\,|\,\sigma^2, \boldsymbol{X}, \boldsymbol{Y})$$

$$= p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y}) \Big/ \underbrace{p(\sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y})}_{\text{not a function of } \boldsymbol{a}} \propto p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y})$$

$$\propto \exp\left[-\frac{1}{2\sigma^2} \sum_{k=0}^{M-1} \|\boldsymbol{r}(i) - \boldsymbol{SB}(i)\boldsymbol{a}\|^2\right]$$

$$\cdot \exp\left[-\tfrac{1}{2}(\boldsymbol{a}-\boldsymbol{a}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{a}-\boldsymbol{a}_0)\right]$$

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{a}^T \underbrace{\left[\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2}\sum_{k=0}^{M-1} \boldsymbol{B}(i)\boldsymbol{S}^T\boldsymbol{SB}(i)\right]}_{\boldsymbol{\Sigma}_*^{-1}}\boldsymbol{a}\right.$$

$$\left.+ \boldsymbol{a}^T \underbrace{\left[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{a}_0 + \frac{1}{\sigma^2}\sum_{k=0}^{M-1} \boldsymbol{B}(i)\boldsymbol{S}^T\boldsymbol{r}(i)\right]}_{\boldsymbol{\Sigma}_*^{-1}\boldsymbol{a}_*}\right\}$$

$$\propto \exp\left[-\tfrac{1}{2}(\boldsymbol{a}-\boldsymbol{a}_*)^T \boldsymbol{\Sigma}_*^{-1}(\boldsymbol{a}-\boldsymbol{a}_*)\right] \sim \mathcal{N}(\boldsymbol{a}_*, \boldsymbol{\Sigma}_*). \quad (54)$$

Derivation of (23):

$$p(\sigma^2\,|\,\boldsymbol{a}, \boldsymbol{X}, \boldsymbol{Y})$$

$$= p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y}) \Big/ \underbrace{p(\boldsymbol{a}, \boldsymbol{X}\,|\,\boldsymbol{Y})}_{\text{not a function of } \sigma^2} \propto p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y})$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{PM/2} \exp\left(-\frac{1}{2\sigma^2} \underbrace{\sum_{k=0}^{M-1} \|\boldsymbol{r}(i) - \boldsymbol{SAx}(i)\|^2}_{s^2}\right)$$

$$\cdot \left(\frac{1}{\sigma^2}\right)^{(\nu_0/2)+1} \exp\left(-\frac{\nu_0\lambda_0}{2\sigma^2}\right)$$

$$= \left(\frac{1}{\sigma^2}\right)^{((\nu_0+PM)/2)+1} \exp\left(-\frac{\nu_0\lambda_0 + s^2}{2\sigma^2}\right)$$

$$\sim \chi^{-2}\left(\nu_0 + PM, \frac{\nu_0\lambda_0 + s^2}{\nu_0 + PM}\right) \quad (55)$$

Derivation of (25):

$$P\left[x_k(i) = +1\,|\,\boldsymbol{a}, \sigma^2, \boldsymbol{X}_{ki}, \boldsymbol{Y}\right]$$

$$= p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y}) \Big/ \underbrace{p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}_{ki}\,|\,\boldsymbol{Y})}_{\text{not a function of } x_k(i)}$$

$$\propto p(\boldsymbol{a}, \sigma^2, \boldsymbol{X}\,|\,\boldsymbol{Y})$$

$$\propto \rho_k(i) \exp\left(-\frac{1}{2\sigma^2} \sum_{l=0}^{M-1} \|\boldsymbol{r}(l) - \boldsymbol{SAx}(l)\|^2\right)$$

$$\propto \rho_k(i) \exp\left(-\frac{1}{2\sigma^2} \|\boldsymbol{r}(i) - \boldsymbol{SAx}(i)\|^2\right) \quad (56)$$

$$\Rightarrow \frac{P\left[x_k(i) = +1\,|\,\boldsymbol{a}, \sigma^2, \boldsymbol{X}_{ki}, \boldsymbol{Y}\right]}{P\left[x_k(i) = -1\,|\,\boldsymbol{a}, \sigma^2, \boldsymbol{X}_{ki}, \boldsymbol{Y}\right]}$$

$$= \frac{\rho_k(i)}{1-\rho_k(i)} \cdot \exp\left\{\frac{1}{2\sigma^2}\left[\left\|\boldsymbol{r}(i) - \boldsymbol{SA}\left(\boldsymbol{x}_k^0(i) - \boldsymbol{1}_k\right)\right\|^2\right.\right.$$

$$\left.\left.- \left\|\boldsymbol{r}(i) - \boldsymbol{SA}\left(\boldsymbol{x}_k^0(i) + \boldsymbol{1}_k\right)\right\|^2\right]\right\}$$

$$= \frac{\rho_k(i)}{1-\rho_k(i)} \cdot \exp\left\{\frac{2}{\sigma^2}(\boldsymbol{SA1}_k)^T\left[\boldsymbol{r}(i) - \boldsymbol{SAx}_k^0(i)\right]\right\}$$

$$= \frac{\rho_k(i)}{1-\rho_k(i)} \cdot \exp\left\{\frac{2A_k}{\sigma^2}\boldsymbol{s}_k^T\left[\boldsymbol{r}(i) - \boldsymbol{SAx}_k^0(i)\right]\right\} \quad (57)$$

[$\boldsymbol{1}_k$ is K-dimensional vector with all-zero entries except for the kth entry, which is 1.]

APPENDIX B

Derivation of (40):

$$p(\boldsymbol{a}\,|\,\sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y})$$

$$= p(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\,|\,\boldsymbol{Y}) \Big/ \underbrace{p(\sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\,|\,\boldsymbol{Y})}_{\text{not a function of } \boldsymbol{a}}$$

$$\propto p(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\,|\,\boldsymbol{Y})$$

$$\propto \exp\left\{-\tfrac{1}{2}\sum_{i=0}^{M-1}[\boldsymbol{r}(i)-\boldsymbol{SB}(i)\boldsymbol{a}]^T \boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i)-\boldsymbol{SB}(i)\boldsymbol{a}]\right\}$$

$$\cdot \exp\left[-\tfrac{1}{2}(\boldsymbol{a}-\boldsymbol{a}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{a}-\boldsymbol{a}_0)\right]$$

$$\propto \exp\left\{ -\tfrac{1}{2} \boldsymbol{a}^T \underbrace{\left[\boldsymbol{\Sigma}_0^{-1} + \sum_{i=0}^{M-1} \boldsymbol{B}(i) \boldsymbol{S}^T \boldsymbol{\Lambda}(i)^{-1} \boldsymbol{S} \boldsymbol{B}(i) \right]}_{\boldsymbol{\Sigma}_*^{-1}} \boldsymbol{a} \right.$$

$$\left. + \boldsymbol{a}^T \underbrace{\left[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{a}_0 + \sum_{i=0}^{M-1} \boldsymbol{B}(i) \boldsymbol{S}^T \boldsymbol{\Lambda}(i)^{-1} \boldsymbol{r}(i) \right]}_{\boldsymbol{\Sigma}_*^{-1}\boldsymbol{a}_*} \right\}$$

$$\propto \exp\left[-\tfrac{1}{2}(\boldsymbol{a}-\boldsymbol{a}_*)^H \boldsymbol{\Sigma}_*^{-1}(\boldsymbol{a}-\boldsymbol{a}_*)\right] \sim \mathcal{N}(\boldsymbol{a}_*, \boldsymbol{\Sigma}_*). \quad (58)$$

Derivation of (43):

$$p\left(\sigma_l^2 \,|\, \boldsymbol{a}, \sigma_{\bar{l}}^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)$$

$$= p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{a}, \sigma_{\bar{l}}^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right)}_{\text{not a function of } \sigma_l^2}$$

$$\propto p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right)$$

$$\propto \left(\frac{1}{\sigma_l^2}\right)^{(1/2)\sum_{i=0}^{M-1} n_l(i)}$$

$$\cdot \exp\left\{ -\frac{1}{2\sigma_l^2} \underbrace{\sum_{i=0}^{M-1}\sum_{j=0}^{P-1}\left[r_j(i) - \boldsymbol{\xi}_j^T \boldsymbol{A}\boldsymbol{x}(i) \right]^2 \mathbf{1}_{\{I_j(i)=l\}}}_{s_l^2} \right\}$$

$$\cdot \left(\frac{1}{\sigma_l^2}\right)^{(\nu_l/2)+1} \exp\left(-\frac{\nu_l \lambda_l}{2\sigma_l^2}\right)$$

$$= \left(\frac{1}{\sigma_l^2}\right)^{(\nu_l/2)+(1/2)\sum_{i=0}^{M-1} n_l(i)+1} \exp\left(-\frac{\nu_l \lambda_l}{2\sigma_l^2}\right)$$

$$\sim \chi^{-2}\left(\left[\nu_l + \sum_{i=0}^{M-1} n_l(i)\right], \frac{\nu_l \lambda_l + s_l^2}{\nu_l + \sum_{i=0}^{M-1} n_l(i)}\right). \quad (59)$$

Derivation of (45):

$$P\left[x_k(i) = +1 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{ki}, \boldsymbol{Y}\right]$$

$$= p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{ki} \,|\, \boldsymbol{Y}\right)}_{\text{not a function of } x_k(i)}$$

$$\propto p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right)$$

$$\propto \rho_k(i) \exp\left\{ -\tfrac{1}{2} \sum_{l=0}^{M-1} [\boldsymbol{r}(l) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(l)]^T \right.$$

$$\left. \cdot \boldsymbol{\Lambda}(l)^{-1}[\boldsymbol{r}(l) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(l)] \right\}$$

$$\propto \rho_k(i) \exp\left\{-\tfrac{1}{2}[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(i)]^T \boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(i)]\right\}$$

$$(60)$$

$$\Rightarrow \frac{P[x_k(i) = +1 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{X}_{ki}, \boldsymbol{Y}]}{P[x_k(i) = -1 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{X}_{ki}, \boldsymbol{Y}]}$$

$$= \frac{\rho_k(i)}{1 - \rho_k(i)} \cdot \exp\Big[\tfrac{1}{2}\{[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}(\boldsymbol{x}(i) - \mathbf{1}_k)]^T$$

$$\cdot \boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}(\boldsymbol{x}(i) - \mathbf{1}_k)] - [\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}(\boldsymbol{x}(i) + \mathbf{1}_k)]^T$$

$$\cdot \boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}(\boldsymbol{x}(i) + \mathbf{1}_k)]\}\Big]$$

$$= \frac{\rho_k(i)}{1 - \rho_k(i)} \cdot \exp\{2(\boldsymbol{S}\boldsymbol{A}\mathbf{1}_k)^T \boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}_k^0(i)]\}$$

$$= \frac{\rho_k(i)}{1 - \rho_k(i)} \cdot \exp\{2A_k \boldsymbol{s}_k^T \boldsymbol{\Lambda}(i)^{-1}[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}_k^0(i)]\}. \quad (61)$$

Derivation of (46):

$$P\left[I_j(i) = l \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{ji}, \boldsymbol{X}, \boldsymbol{Y}\right]$$

$$= p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{ji}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right)}_{\text{not a function of } I_j(i)}$$

$$\propto p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right) \propto P[I_j(i) = l \,|\, \epsilon]$$

$$\cdot \frac{1}{\sqrt{\sigma_l^2}} \exp\left\{ -\frac{1}{2\sigma_l^2}\left[r_j(i) - \boldsymbol{\xi}_j^T \boldsymbol{A}\boldsymbol{x}(i) \right]^2 \right\} \quad (62)$$

$$\Rightarrow \frac{P[I_j(i) = 1 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{ji}, \boldsymbol{X}, \boldsymbol{Y}]}{P[I_j(i) = 2 \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{ji}, \boldsymbol{X}, \boldsymbol{Y}]}$$

$$= \frac{1-\epsilon}{\epsilon} \cdot \sqrt{\frac{\sigma_2^2}{\sigma_1^2}} \cdot \exp\left\{ \tfrac{1}{2}\left[r_j(i) - \boldsymbol{\xi}_j^T \boldsymbol{A}\boldsymbol{x}(i) \right]^2 \right.$$

$$\left. \cdot \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) \right\}. \quad (63)$$

Derivation of (47):

$$p\left(\epsilon \,|\, \boldsymbol{a}, \sigma_1^2, \sigma_2^2, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)$$

$$= p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right)}_{\text{not a function of } \epsilon}$$

$$\propto p\left(\boldsymbol{a}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X} \,|\, \boldsymbol{Y}\right) \propto p(\epsilon)p(\boldsymbol{I} \,|\, \epsilon)$$

$$\propto \epsilon^{a_0-1}(1-\epsilon)^{b_0-1} \cdot \epsilon^{\sum_{i=0}^{M-1} n_2(i)}(1-\epsilon)^{\sum_{i=0}^{M-1} n_1(i)}$$

$$\sim \text{Beta}\left(a_0 + \sum_{i=0}^{M-1} n_2(i), \, b_0 + \sum_{i=0}^{M-1} n_1(i)\right). \quad (64)$$

ACKNOWLEDGMENT

REFERENCES

[1] B. Aazhang and H. V. Poor, "Peformance of DS/SSMA communications in impulsive channels—Part I: Linear correlation receivers," *IEEE Trans. Commun.*, vol. COMM-35, pp. 1179–1187, Nov. 1987.

[2] ——, "Peformance of DS/SSMA communications in impulsive channels—Part II: Hard-limiting correlation receivers," *IEEE Trans. Commun.*, vol. COMM-36, pp. 88–96, Jan. 1988.

[3] ——, "An analysis of nonlinear direct-sequence correlators," *IEEE Trans. Commun.*, vol. COMM-37, pp. 723–731, July 1989.

[4] S. F. Arnold, "Gibbs sampling," in *Handbook of Statistics*, C. R. Rao, Ed. New York: Elsevier, 1993, vol. 9, pp. 599–625.

[5] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag, 1985.

[6] C. Berrou and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo codes," *IEEE Trans. Commun.*, vol. COMM-44, Oct. 1996.

[7] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correction coding and decoding: Turbo codes," in *Proc. 1993 Int. Conf. Commun.*, 1993, pp. 1064–1070.

[8] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 991–1001, Sept. 1993.

[9] T. K. Blankenship, D. M. Krizman, and T. S. Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," in *Proc. IEEE Veh. Technol. Conf.*, 1997, pp. 1942–1946.

[10] G. E. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, 1973.

[11] P. L. Brockett, M. Hinich, and G. R. Wilson, "Nonlinear and non-Gaussian ocean noise," *J. Acoust. Soc. Amer.*, vol. 82, pp. 1286–1399, 1987.

[12] G. Casella and E. I. George, "Explaining the Gibbs sampler," *Amer. Stat.*, vol. 46, pp. 167–174, 1992.

[13] K. S. Chan, "Asymptotic behavior of the Gibbs sampler," *J. Amer. Stat. Assoc.*, vol. 88, pp. 320–326, 1993.

[14] U. Fawer and B. Aazhang, "A multiuser receiver for code-division multiple-access communications over multipath channels," *IEEE Trans. Commun.*, vol. COMM-43, no. 2/3/4, pp. 1556–1565, Feb./Mar./Apr. 1995.

[15] A. E. Gelfand and A. F. W. Smith, "Sampling-based approaches to calculating marginal densities," *J. Amer. Stat. Assoc.*, vol. 85, pp. 398–409, 1990.

[16] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1995.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[18] T. R. Giallorenzi and S. G. Wilson, "Multiuser ML sequence estimator for convolutional coded asynchronous DS-CDMA systems," *IEEE Trans. Commun.*, vol. COMM-44, pp. 997–1008, Aug. 1996.

[19] ——, "Suboptimum multiuser receivers for convolutionally coded asynchrnous DS-CDMA systems," *IEEE Trans. Commun.*, vol. COMM-44, pp. 1183–1196, Sept. 1996.

[20] P. J. Green, "Revisible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1985.

[21] J. Hagenauer, "The Turbo principle: Tutorial introduction and state of the art," in *Proc. Int. Symp. Turbo Codes Related Topics*, Brest, France, Sept. 1997, pp. 1–11.

[22] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[23] R. A. Iltis, "An EKF-based joint estimator for interference, multipath and code delay in a DS spread-spectrum receiver," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1288–1299, Feb./Mar./Apr. 1994.

[24] R. A. Iltis and L. Mailaender, "An adaptive multiuser detector with joint amplitude and delay estimation," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 774–785, June 1994.

[25] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag, 1998.

[26] J. S. Liu, A. Kong, and W. H. Wong, "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes," *Biometrika*, vol. 81, pp. 27–40, 1994.

[27] N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.

[28] D. Middleton, "Man-made noise in urban environments and transportation systems: Models and measurement," *IEEE Trans. Commun.*, vol. COMM-21, pp. 1232–1241, Nov. 1973.

[29] ——, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, pp. 106–127, 1977.

[30] ——, "Channel modeling and threshold signal processing in underwater acoustics: An analytical overview," *IEEE J. Ocean. Eng.*, vol. OE-12, pp. 4–28, 1987.

[31] ——, "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1122–1129, May 1999.

[32] D. Middleton and A. D. Spaulding, "Elements of weak signal detection in non-Gaussian noise," in *Adv. Stat. Signal Process. Vol. 2: Signal Detection*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI, 1993.

[33] S. Y. Miller, "Detection and estimation in multiple-access channels," Ph.D. thesis, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, 1989.

[34] M. Moher, "An iterative multiuser decoder for near-capacity communications," *IEEE Trans. Commun.*, vol. COMM-46, pp. 870–880, July 1998.

[35] T. K. Moon, Z. Xie, C. K. Rushforth, and R. T. Short, "Parameter estimation in a multiuser communication system," *IEEE Trans. Commun.*, vol. 42, pp. 2553–2560, Aug. 1994.

[36] H. V. Poor, "On parameter estimation in DS/CDMA formats," in *Adv. Commun. Signal Process.*, W. A. Porter and S. C. Kak, Eds. New York: Springer-Verlag, 1989.

[37] H. V. Poor and M. Tanda, "An analysis of some multiuser detectors in impulsive noise," in *Proc. 16th GRETSI Symp. Signal Image Process.*, Grenoble, France, Sept. 1997.

[38] M. C. Reed, C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer, "Iterative multiuser detection for CDMA with FEC: Near single user performance," *IEEE Trans. Commun.*, vol. 46, pp. 1693–1699, Dec. 1998.

[39] C. P. Robert, *The Bayesian Choice: A Decision-Theoretic Motivation*. New York: Springer-Verlag, 1994.

[40] T. J. Rothenberg, "The Bayesian approach and alternatives in econometrics," in *Studies in Bayesian Econometrics and Statistics, Vol. 1*, S. Fienberg and A. Zellners, Eds. Amsterdam, The Netherlands: North-Holland:, 1977, pp. 55–75.

[41] M. J. Schervish and B. P. Carlin, "On the convergence of successive substitution sampling," *J. Comput.*, vol. 1, pp. 111–127, 1992.

[42] Y. Steinberg and H. V. Poor, "Sequential amplitude estimation in multiuser communications," *IEEE Trans. Inform. Theory*, vol. 40, pp. 11–20, Jan. 1994.

[43] E. Strom and S. Miller, "Asynchronous DS-CDMA systems: Low complexity near–far resistant receiver and parameter estimation," Tech. Rep., Dept. Elect. Eng., Univ. Florida, Gainesville, Jan. 1994.

[44] M. A. Tanner, *Tools for Statistics Inference*. New York: Springer-Verlag, 1991.

[45] M. A. Tanner and W. H. Wong, "The calculation of posterior distribution by data augmentation (with discussion)," *J. Amer. Statist. Assoc.*, vol. 82, pp. 528–550, 1987.

[46] S. Verdú, "Optimal multiuser signal detection," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Illinois, Urbana, 1984.

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

[48] ——, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.

[49] ——, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[50] J. von Neumann, "Various techniques used in connection with random digit," *Natl. Bur. Stand. Appl. Math. Ser.*, vol. 12, pp. 36–38, 1951.

[51] X. Wang and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, July 1999.

[52] ——, "Robust multiuser detection in non-Gaussian channels," *IEEE Trans. Signal Processing*, vol. 47, pp. 289–305, Feb. 1999.

[53] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[54] W. H. Wong and F. Liang, "Dynamic importance weighting in Monte Carlo and optimization," *Proc. Nat. Acad. Sci.*, vol. 94, pp. 14 220–14 224, 1997.

[55] Z. Xie, C. K. Rushforth, R. T. Short, and T. K. Moon, "Joint signal detection and parameter estimation in multiuser communications," *IEEE Trans. Commun.*, vol. 41, pp. 1208–1216, Aug. 1993.

[56] S. M. Zabin and H. V. Poor, "Efficient estimation of the class A parameters via the EM algorithm," *IEEE Trans. Inform. Theory*, vol. 37, pp. 60–72, Jan. 1991.

Xiaodong Wang (M'98) received the B.S. degree in electrical engineering and applied mathematics (with the highest honor) from Shanghai Jiao Tong University, Shanghai, China, in 1992, the M.S. degree in electrical and computer engineering form Purdue University, West Lafayette, IN, in 1995, and the Ph.D. degree in electrical engineering from Princeton Univesity, Princeton, NJ, in 1998.

In July 1998, he joined the Department of Electrical Engineering, Texas A&M University, College Station, as an Assistant Professor. His research interests fall in the general areas of computing, signal processing, and communications. He has worked in the areas of digital communications, digital signal processing, parallel and distributed computing, nanoelectronics, and quantum computing. His current research interests include multiuser communications theory and advanced signal processing for wireless communications. He worked at the AT&T Bell Laboratories-Research, Red Bank, NJ, during the summer of 1997.

Dr. Wang is a member of the American Association for the Advancement of Science. He received the NSF CAREER Award in 1999.

Rong Chen received the B.S. degree in mathematics from Peking University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in statistics from Carnegie Mellon University, Pittsburgh, PA, in 1987 and 1990, respectively.

Since September 1990, he has been with Texas A&M University, College Station, where he is currently as Associate Professor of statistics. His current research interests include nonlinear/nonparametric time series and Monte Carlo methods for nonlinear non-Gaussian dynamic systems.

Dr. Chen is a member of the American Statistical Association and the Institute of Mathematical Statistics.