Adaptive Bayesian Multiuser Detection *

Xiaodong Wang
Department of Electrical Engineering
Texas A& M University
College Station, TX 77843

Rong Chen
Department of Statistics
Texas A&M University
College Station, TX 77843

Abstract

We consider the problem of simultaneous parameter estimation and data restoration in a synchronous CDMA system, in the presence additive Gaussian white noise with unknown parameters. 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 unknowns, 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 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 as *adaptive Turbo multiuser detector.*

1 Introduction

The theme of this paper is to treat two related problems in multiuser detection under a general Bayesian framework. These problems are: (i) Optimal multiuser detection in the presence of

*This work was supported in part by the Interdisciplinary Research Initiatives Program, Texas A&M University. X. Wang was supported in part by the U.S. National Science Foundation under grant CAREER CCR 9875314. R. Chen was supported in part by the U.S. National Science Foundation under grant DMS 9626113.

unknown channel parameters; (ii) Multiuser detection for coded CDMA systems. We first provide a perspective on the related works in these two areas.

Optimal multiuser detection with unknown parameters: The optimal multiuser detection algorithms with known channel parameters, that is, the multiuser maximum-likelihood sequence detector (MLSD), and the multiuser minimum *a posteriori* probability (MAP) detector, were first investigated in [26, 27] (cf.[29]). The analysis of the computational complexity and the proof that the optimal multiuser detection problem is combinatorially hard appeared in [26, 28]. When the channel parameters (e.g., received amplitudes, 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 [19], where a solution based on the expectation-maximization (EM) algorithm is derived. In [22], 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 [34], a tree-search algorithm is given for joint data detection and amplitude estimation. Other works concerning multiuser detection with unknown channel parameters include [6, 13, 14, 16, 18, 23].

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 [9], the optimal decoding scheme for convolutionally coded CDMA system is studied, which is shown to have a prohibitive computational complexity. In [10], some low-complexity receivers which perform multiuser symbol detection and decoding either separately or jointly are studied. In [17, 20, 31], 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. We consider Bayesian inference of all unknown quantities (e.g., received amplitudes, data symbols, noise variance) 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 2, the system under study is described. In Section 3, some background material on the Gibbs sampler is provided. The problems of adaptive Bayesian multiuser detection in synchronous CDMA channels are treated in Section 4 and Section 5 respectively. In Section 6, 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 7. Simulation results are provided in Section 8. Finally, Section 9 contains the conclusions.

2 System Description

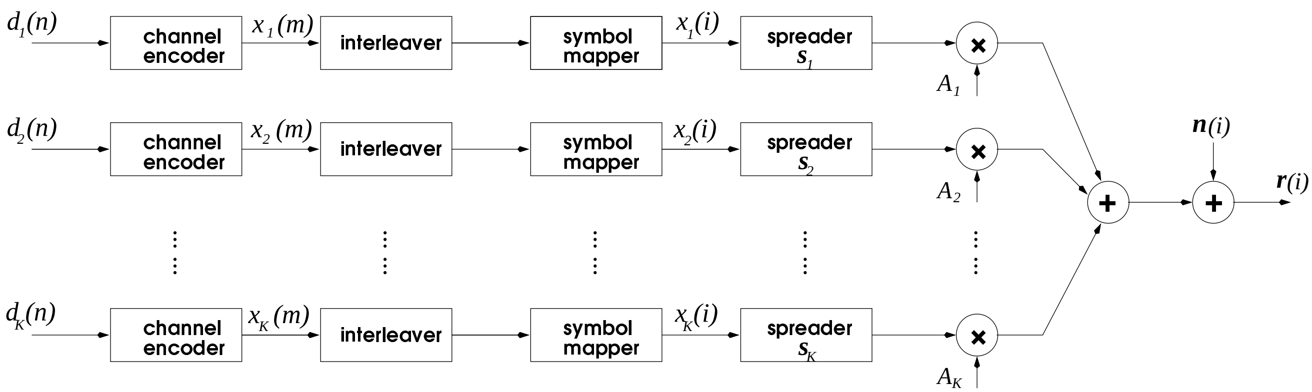

Figure 1: A coded synchronous CDMA communication system.

We consider a coded synchronous CDMA system with K users, employing normalized modulation waveforms $\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K$, and signaling through a channel with additive white noise. The block diagram of the transmitter end of such a system is shown in Figure 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 $\boldsymbol{s}_k$, and transmitted through the channel. The received signal is the superposition of the K users' transmitted signals plus the ambient noise, given by

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

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

$$\boldsymbol{s}_k = [\beta_{k,0}\ \beta_{k,1}\ \cdots\ \beta_{k,P-1}]^T, \quad \beta_{k,j} \in \{+1, -1\}, \tag{2}$$

where P is the spreading factor. Define the following *a priori* symbol probabilities

$$\rho_k(i) \ \stackrel{\triangle}{=}\ P[x_k(i) = +1], \quad i = 0, \cdots, M-1;\ k = 1, \cdots, K. \tag{3}$$

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 $\{\boldsymbol{n}(i)\}$ 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 $\boldsymbol{n}(i)$ is assumed to consist of i.i.d. samples $\{n_j(i)\}_{j=0}^{P-1}$. The noise $n_j(i)$ is assumed to have a Gaussian distribution, i.e.,

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

where σ^2 is the variance of the noise.

Denote $\boldsymbol{Y} \stackrel{\triangle}{=} \{\boldsymbol{r}(0), \boldsymbol{r}(1), \cdots, \boldsymbol{r}(M-1)\}$. In Sections 4 and 5, we consider the problem of estimating the *a posteriori* probabilities of the transmitted symbols

$$P\left[x_k(i) = +1 \mid \boldsymbol{Y}\right], \quad i = 0, \cdots, M-1;\ k = 1, \cdots, K, \tag{5}$$

based on the received signals $\boldsymbol{Y}$ 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; 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 Figure 1, which will be discussed in Section 6.

3 The 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 [24]. 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* [7], which is described next.

Let $\boldsymbol{\theta} = [\theta_1 \ \cdots \ \theta_d]^T$ be a vector of unknown parameters and let $\boldsymbol{Y}$ 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 $\boldsymbol{Y}$, i.e., $p(\theta_j \mid \boldsymbol{Y})$, $1 \leq j \leq d$. Direct evaluation involves integrating out the rest of the parameters from the joint *a posteriori* density, i.e.,

$$p(\theta_j \mid \boldsymbol{Y}) \quad = \quad \int\int\cdots\int p(\boldsymbol{\theta} \mid \boldsymbol{Y})\, \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{j-1}\mathrm{d}\theta_{j+1} \cdots \mathrm{d}\theta_d. \tag{6}$$

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 $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$, and then to estimate any marginal distribution using these samples. Given the initial values $\boldsymbol{\theta}^{(0)} = [\theta_1^{(0)} \ \cdots \ \theta_d^{(0)}]^T$, this algorithm iterates the following loop:

- Draw sample $\theta_1^{(n+1)}$ from $p\left(\theta_1 \mid \theta_2^{(n)}, \cdots, \theta_d^{(n)}, \boldsymbol{Y}\right)$;
- Draw sample $\theta_2^{(n+1)}$ from $p\left(\theta_2 \mid \theta_1^{(n+1)}, \theta_3^{(n)}, \cdots, \theta_d^{(n)}, \boldsymbol{Y}\right)$;

  ⋮
- Draw sample $\theta_d^{(n+1)}$ from $p\left(\theta_d \mid \theta_1^{(n+1)}, \cdots, \theta_{d-1}^{(n+1)}, \boldsymbol{Y}\right)$.

Under regularity conditions, the vectors $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(n)} \cdots$ are a realization of a homogeneous Markov chain with the transition kernel from state $\boldsymbol{\theta}'$ to state $\boldsymbol{\theta}$, given by

$$K\left(\boldsymbol{\theta}', \boldsymbol{\theta}\right) \quad = \quad p\left(\theta_1 \mid \theta_2', \cdots, \theta_d', \boldsymbol{Y}\right)\, p\left(\theta_2 \mid \theta_1, \theta_3', \cdots, \theta_d', \boldsymbol{Y}\right) \cdots p\left(\theta_d \mid \theta_1, \cdots, \theta_{d-1}, \boldsymbol{Y}\right). \tag{7}$$

The convergence behavior of the Gibbs sampler is investigated in [3, 7, 8, 15, 21, 25] and general conditions are given for the following two results:

- The distribution of $\boldsymbol{\theta}^{(n)}$ converges geometrically to $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$, as $n \longrightarrow \infty$.
- $\frac{1}{N}\sum_{n=1}^{N} f\left(\boldsymbol{\theta}^{(n)}\right) \stackrel{\text{a.s.}}{\longrightarrow} \int f(\boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{Y})\, \mathrm{d}\boldsymbol{\theta}$, 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 "burning-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 [24].

4 Adaptive Bayesian Multiuser Detector

In this section, we consider the problem of computing the *a posteriori* probabilities in (5) under the assumption that the ambient noise distribution is Gaussian. That is, the pdf of $\boldsymbol{n}(i)$ in (1) is given by

$$p(\boldsymbol{n}(i)) \;=\; \frac{1}{(2\pi\sigma^2)^{\frac{P}{2}}} \exp\left(-\frac{\|\boldsymbol{n}(i)\|^2}{2\sigma^2}\right). \tag{8}$$

Denote

$$\begin{aligned}
\boldsymbol{x}(i) &\triangleq [x_1(i)\; x_2(i)\; \cdots\; x_K(i)]^T, \quad i = 0, 1, \cdots, M-1, \\
\boldsymbol{B}(i) &\triangleq \mathsf{diag}(x_1(i), x_2(i), \cdots, x_K(i)), \quad i = 0, 1, \cdots, M-1, \\
\boldsymbol{X} &\triangleq [\boldsymbol{x}(0)\; \boldsymbol{x}(1)\; \cdots\; \boldsymbol{x}(M-1)], \\
\boldsymbol{Y} &\triangleq [\boldsymbol{r}(0)\; \boldsymbol{r}(1)\; \cdots\; \boldsymbol{r}(M-1)], \\
\boldsymbol{a} &\triangleq [A_1\; A_2\; \cdots\; A_K]^T, \\
\boldsymbol{A} &\triangleq \mathsf{diag}(A_1, A_2, \cdots, A_K), \\
\boldsymbol{S} &\triangleq [\boldsymbol{s}_1\; \boldsymbol{s}_2\; \cdots\; \boldsymbol{s}_K].
\end{aligned}$$

Then (1) can be written as

$$\boldsymbol{r}(i) \;=\; \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(i) + \boldsymbol{n}(i) \tag{9}$$

$$\;=\; \boldsymbol{S}\boldsymbol{B}(i)\boldsymbol{a} + \boldsymbol{n}(i), \qquad i = 0, 1, \cdots, M-1. \tag{10}$$

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

4.1 Bayesian Inference

Assume that the unknown quantities $\boldsymbol{a}$, σ^2 and $\boldsymbol{X}$ are independent of each other and have prior distributions $p(\boldsymbol{a})$, $p(\sigma^2)$ and $p(\boldsymbol{X})$, respectively. Since $\{\boldsymbol{n}(i)\}_{i=0}^{M-1}$ is a sequence of independent Gaussian vectors, using (8) and (9), the joint posterior distribution of these unknown quantities $(\boldsymbol{a}, \sigma^2, \boldsymbol{X})$ based on the received signal $\boldsymbol{Y}$ takes the form of

$$p\left(\boldsymbol{a}, \sigma^2, \boldsymbol{X} \mid \boldsymbol{Y}\right) \;\propto\; p\left(\boldsymbol{Y} \mid \boldsymbol{a}, \sigma^2, \boldsymbol{X}\right) p\left(\boldsymbol{a}\right) p\left(\sigma^2\right) p\left(\boldsymbol{X}\right)$$

$$\propto\; \left(\frac{1}{\sigma^2}\right)^{\frac{PM}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{M-1}\|\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(i)\|^2\right) p\left(\boldsymbol{a}\right) p\left(\sigma^2\right) p\left(\boldsymbol{X}\right). \tag{11}$$

The *a posteriori* probabilities (5) of the transmitted symbols can then be calculated from the joint posterior distribution (11) according to

$$P\left[x_k(i) = +1 \mid \boldsymbol{Y}\right] = \sum_{\boldsymbol{X}:\, x_k(i)=+1} p\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) = \sum_{\boldsymbol{X}:\, x_k(i)=+1} \int p\left(\boldsymbol{a}, \sigma^2, \boldsymbol{X} \mid \boldsymbol{Y}\right) \mathrm{d}\boldsymbol{a}\, \mathrm{d}\sigma^2. \tag{12}$$

Clearly the computation in (12) involves 2^{M-1} multi-dimensional integrals, which is certainly infeasible for any practical implementations. To avoid the direct evaluation of the Bayesian estimate (12), we resort to the Gibbs sampler discussed in Section 3. The basic idea is to generate ergodic random samples $\left\{\boldsymbol{a}^{(n)}, \sigma^{2(n)}, \boldsymbol{X}^{(n)} : n = 0, 1, \cdots\right\}$ from the posterior distribution (11), and then to average $\{x_k(i)^{(n)} : n = 0, 1, \cdots\}$ to obtain an approximation of the *a posteriori* probabilities in (12).

4.2 Prior Distributions

In principle, prior distributions are used to incorporate the prior knowledge about the unknown parameters, and less restrictive (or less informative) priors should be employed when such knowledge is limited. Computational complexity is another consideration that affects the selection. Conjugate priors are usually used to obtain simple analytical forms for the resulting posterior distributions. To make the Gibbs sampler more computationally efficient, the priors should also be chosen such that the conditional posterior distributions are easy to simulate. We next specify the prior distributions $p(\boldsymbol{a})$, $p\left(\sigma^2\right)$ and $p(\boldsymbol{X})$.

For the unknown amplitudes $\boldsymbol{a}$, a complex Gaussian prior distribution is assumed,

$$p(\boldsymbol{a}) \quad \sim \quad \mathcal{N}\left(\boldsymbol{a}_0, \boldsymbol{\Sigma}_0\right). \tag{13}$$

Note that large value of $\boldsymbol{\Sigma}_0$ corresponds to the less informative prior. For the noise variance σ^2, an inverse chi-square prior distribution is assumed,

$$p\left(\sigma^2\right) = \frac{\left(\frac{\nu_0\lambda_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{\nu_0}{2}+1} \exp\left(-\frac{\nu_0\lambda_0}{2\sigma^2}\right) \quad \sim \quad \chi^{-2}(\nu_0,\, \lambda_0), \tag{14}$$

$$\text{or} \qquad \frac{\nu_0\lambda_0}{\sigma^2} \quad \sim \quad \chi^2(\nu_0). \tag{15}$$

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 $p(\boldsymbol{X})$ can be expressed in terms of the prior symbol probabilities defined in (3) as

$$p(\boldsymbol{X}) = \prod_{i=0}^{M-1} \prod_{k=1}^{K} \rho_k(i)^{\delta_{ki}} [1-\rho_k(i)]^{1-\delta_{ki}}, \tag{16}$$

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

4.3 Conditional Posterior Distributions

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

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

$$p\left(\boldsymbol{a} \mid \sigma^2, \boldsymbol{X}, \boldsymbol{Y}\right) \sim \mathcal{N}\left(\boldsymbol{a}_*, \boldsymbol{\Sigma}_*\right), \tag{17}$$

$$\text{with} \quad \boldsymbol{\Sigma}_*^{-1} \stackrel{\triangle}{=} \boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \boldsymbol{B}(i) \boldsymbol{R} \boldsymbol{B}(i), \tag{18}$$

$$\text{and} \quad \boldsymbol{a}_* \stackrel{\triangle}{=} \boldsymbol{\Sigma}_* \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{a}_0 + \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \boldsymbol{B}(i) \boldsymbol{S}^T \boldsymbol{r}(i) \right), \tag{19}$$

where in (18) $\boldsymbol{R} \stackrel{\triangle}{=} \boldsymbol{S}^T \boldsymbol{S}$.

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

$$p\left(\sigma^2 \mid \boldsymbol{a}, \boldsymbol{X}, \boldsymbol{Y}\right) \sim \chi^{-2}\left(\nu_0 + PM, \frac{\nu_0\lambda_0 + s^2}{\nu_0 + PM}\right), \tag{20}$$

$$\text{or} \quad \frac{(\nu_0\lambda_0 + s^2)}{\sigma^2} \sim \chi^2\left(\nu_0 + PM\right), \tag{21}$$

$$\text{with} \quad s^2 \stackrel{\triangle}{=} \sum_{i=0}^{M-1} \|\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}(i)\|^2. \tag{22}$$

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 $x_k(i)$.]

$$\frac{P\left[x_k(i) = +1 \mid \boldsymbol{a}, \sigma^2, \boldsymbol{X}_{ki}, \boldsymbol{Y}\right]}{P\left[x_k(i) = -1 \mid \boldsymbol{a}, \sigma^2, \boldsymbol{X}_{ki}, \boldsymbol{Y}\right]} = \frac{\rho_k(i)}{1-\rho_k(i)} \cdot \exp\left\{ \frac{2A_k}{\sigma^2} \boldsymbol{s}_k^T \left[\boldsymbol{r}(i) - \boldsymbol{S}\boldsymbol{A}\boldsymbol{x}_k^0(i)\right] \right\},$$
$$k = 1, \cdots, K; \quad i = 0, \cdots, M-1. \tag{23}$$

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

4.4 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 $\left\{\boldsymbol{a}^{(0)},\ \sigma^{2(0)},\ \boldsymbol{X}^{(0)}\right\}$ drawn from their prior distributions, and for $n = 1, 2, \cdots$

1. Draw $\boldsymbol{a}^{(n)}$ from $p\left(\boldsymbol{a} \mid \sigma^{2(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y}\right)$ given by (17).
2. Draw $\sigma^{2(n)}$ from $p\left(\sigma^2 \mid \boldsymbol{a}^{(n)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y}\right)$ given by (21).
3. For $i = 0,\ 1,\ \cdots,\ M-1$

   For $k = 1,\ 2,\ \cdots,\ K$

   Draw $x_k(i)^{(n)}$ from $P\left[x_k(i) \mid \boldsymbol{a}^{(n)}, \sigma^{2(n)}, \boldsymbol{X}_{ki}^{(n)}, \boldsymbol{Y}\right]$ given by (23),

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

Since the amplitudes A_k's are positive, in the first step we adopt the *constrained* Gibbs sampler [4, 5]. To draw samples of $\boldsymbol{a}$ that satisfy this condition, the so-called *rejection method* [30] can be used. For instance, after a sample is drawn from (17), check to see if the constraints $A_k > 0$, $k = 1, \cdots, K$, are satisfied; if not, the sample is rejected and a new sample is drawn from (17); 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 (12) are approximated as

$$P\left[x_k(i) = +1 \mid \boldsymbol{Y}\right] \ \cong\ \frac{1}{N} \sum_{n=n_0+1}^{n_0+N} \delta_{ki}^{(n)}, \tag{24}$$

where $\delta_{ki}^{(n)}$ is the indicator such that $\delta_{ki}^{(n)} = 1$ if $x_k^{(n)} = +1$ and $\delta_{ki}^{(n)} = 0$ if $x_k^{(n)} = -1$. Furthermore, if desired, the estimates of the amplitudes $\boldsymbol{a}$ and the noise variance σ^2 can also be obtained from the corresponding sample means

$$E\left\{\boldsymbol{a} \mid \boldsymbol{Y}\right\} \ \cong\ \frac{1}{N} \sum_{n=n_0+1}^{n_0+N} \boldsymbol{a}^{(n)}, \tag{25}$$

$$\text{and} \quad E\left\{\sigma^2 \mid \boldsymbol{Y}\right\} \ \cong\ \frac{1}{N} \sum_{n=n_0+1}^{n_0+N} \sigma^{2(n)}. \tag{26}$$

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

$$\text{Cov}\{\boldsymbol{a} \mid \boldsymbol{Y}\} \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N}\left[\boldsymbol{a}^{(n)}\right]\left[\boldsymbol{a}^{(n)}\right]^H - \frac{1}{N^2}\left[\sum_{n=n_0+1}^{n_0+N}\boldsymbol{a}^{(n)}\right]\left[\sum_{n=n_0+1}^{n_0+N}\boldsymbol{a}^{(n)}\right]^H, \quad (27)$$

$$\text{and} \quad \text{Var}\left\{\sigma^2 \mid \boldsymbol{Y}\right\} \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N}\left[\sigma^{2(n)}\right]^2 - \frac{1}{N^2}\left[\sum_{n=n_0+1}^{n_0+N}\sigma^{2(n)}\right]^2. \quad (28)$$

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 is $O(K^3 + KM)$, i.e., it has a term which is cubic with respect to the number of users K [due to the inversion of the matrix in (18)], and a term which is *linear* with respect to the symbol block size M [as opposed to *exponential* as in the direct implementation of the Bayesian symbol estimate (12).].

5 Iterative Joint Multiuser Detection and Decoding – Adaptive Turbo Multiuser Detection

Recently iterative ("Turbo") processing techniques have received considerable attention followed by the discovery of the powerful Turbo codes [1, 2]. 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 [12]. 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 [31], 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 adaptive Turbo multiuser receivers for synchronous CDMA channels, with unknown amplitudes and noise parameters.

The iterative (Turbo) receiver structure is shown in Figure 2. It consists of two stages: the adaptive multiuser detector developed in the previous sections, followed by a soft-input soft-output

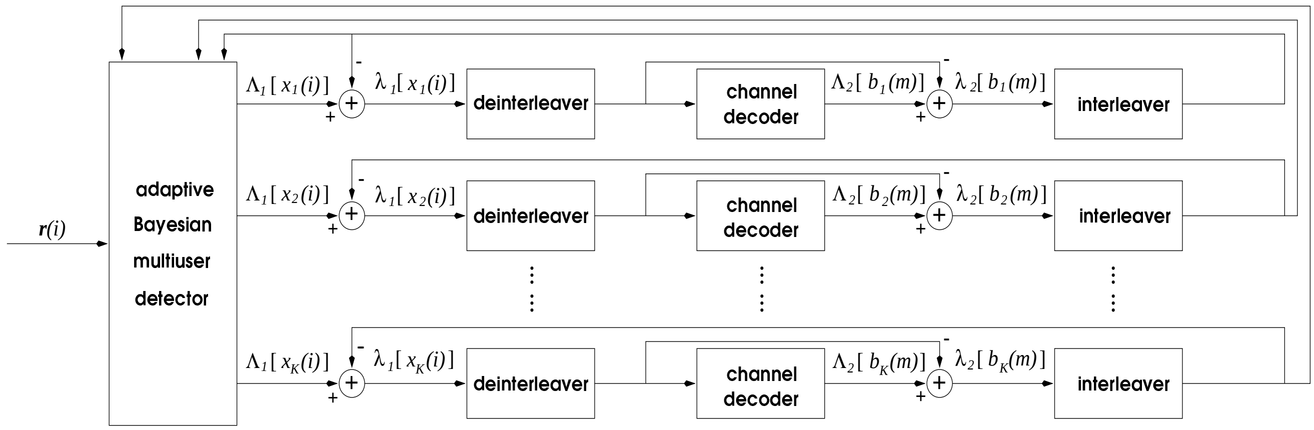

Figure 2: Iterative processing for joint Bayesian multiuser detection and decoding – adaptive Turbo multiuser detection.

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 $\{P[x_k(i) = +1 \mid \boldsymbol{Y}]\}_{k=1;i=0}^{K;M-1}$. Based on these, we first compute the *a posteriori* log-likelihood ratios (LLR's) of a transmitted "+1" symbol and a transmitted "−1" symbol,

$$\Lambda_1[x_k(i)] \triangleq \log \frac{P[x_k(i) = +1 \mid \boldsymbol{Y}]}{P[x_k(i) = -1 \mid \boldsymbol{Y}]}, \quad k = 1, \cdots, K; \quad i = 0, \cdots, M-1. \tag{29}$$

Using the Bayes' rule, (29) can be written as

$$\Lambda_1[x_k(i)] = \underbrace{\log \frac{p[\boldsymbol{Y} \mid x_k(i) = +1]}{p[\boldsymbol{Y} \mid x_k(i) = -1]}}_{\lambda_1[x_k(i)]} + \underbrace{\log \frac{P[x_k(i) = +1]}{P[x_k(i) = -1]}}_{\lambda_2^p[x_k(i)]}, \tag{30}$$

where the second term in (30), 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 (30), denoted by $\lambda_1[x_k(i)]$, represents the *extrinsic* information delivered by the adaptive Bayesian multiuser detector, based on the received signals $\boldsymbol{Y}$, 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 [31],

$$\begin{aligned}\Lambda_2[x_k(m)] &\triangleq \log \frac{P[x_k(m)=+1 \mid \{\lambda_1^p[x_k(i)]\}_{k=1;i=0}^{K;M-1}; \text{decoding}]}{P[x_k(m)=-1 \mid \{\lambda_1^p[x_k(i)]\}_{k=1;i=0}^{K;M-1}; \text{decoding}]} \\ &= \lambda_2[x_k(m)] + \lambda_1^p[x_k(m)]. \end{aligned} \tag{31}$$

It is seen from (31) 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 (5), 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,

$$\begin{aligned}\rho_k(i) &\triangleq P[x_k(i)=+1] = \frac{\exp\left(\lambda_2^p[x_k(i)]\right)}{1+\exp\left(\lambda_2^p[x_k(i)\right)} \\ &= \frac{1}{2}\left[1+\tanh\left(\frac{1}{2}\lambda_2^p[x_k(i)]\right)\right]. \end{aligned} \tag{32}$$

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. But subsequently since they use the same information indirectly, they will become more and more correlated and finally the improvement through the iterations will diminish.

6 Discussions

Decoder-assisted convergence assessment

Detecting convergence in the Gibbs sampler is usually done in some *ad hoc* way. Some methods can be found in [24]. 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 threshhold, then we decide 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 amount 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.

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 the code bits are *not* interleaved. Then one signal frame of M symbols contains $J = fracML$ code words, with the j-th code word given by

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

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 symbols $x_k(i)$ once a time according to (23), 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 a valid code word for any block code [32]). 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

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

$$k = 1, \cdots, K; \quad j = 0, 1, \cdots, \frac{M}{L} - 1, \tag{33}$$

where $\boldsymbol{X}_{kj}$ denotes the set containing all elements of $\boldsymbol{X}$ except for $\underline{x}_k(j)$; $\rho_{kj}(\underline{u}) \stackrel{\triangle}{=} P[\underline{x}_k(j) = \underline{u}]$; and $\boldsymbol{x}_k^0(i) \stackrel{\triangle}{=} [x_1(i), \cdots, x_{k-1}(i), 0, x_{k+1}(i), \cdots, x_K(i)]^T$. 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 can be drawn. This will be demonstrated by simulation examples in the next section.

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 [19]. The major advantage of the Gibbs sampling technique proposed here 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 in the likelihood surface. The EM method performs well only if the initial estimates of the channel and symbols are close enough 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 convegence rate crucially depends on the "energy gap" on the joint posterior density surface. Many modification of the Gibbs sampler have been developed to combat the "large energy gap" situation. For example, see [11, 33].

7 Simulations

In this section, we provide some simulation examples to illustrate the performance of the adaptive Bayesian mulituser detectors developed in this paper. We consider a 5-user ($K = 5$) synchronous CDMA channel with processing gain $P = 10$. The user spreading waveform matrix $\boldsymbol{S}$ and the corresponding correlation matrix $\boldsymbol{R}$ are given respectively by

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

Convergence Behavior of the Gibbs Multiuser Detectors

We first illustrate the performance of the proposed adaptive Bayesian multiuser detector in Gaussian ambient noise. In Figure 3, the convergence behavior of the Gibbs multiuser detector is illustrated for noise level $\sigma^2 = -2$dB. The first 100 samples drawn by the Gibbs sampler for the user amplitudes $(A_1, A_2, A_3, A_4, A_5)$ and the noise variance σ^2 are shown. The corresponding true

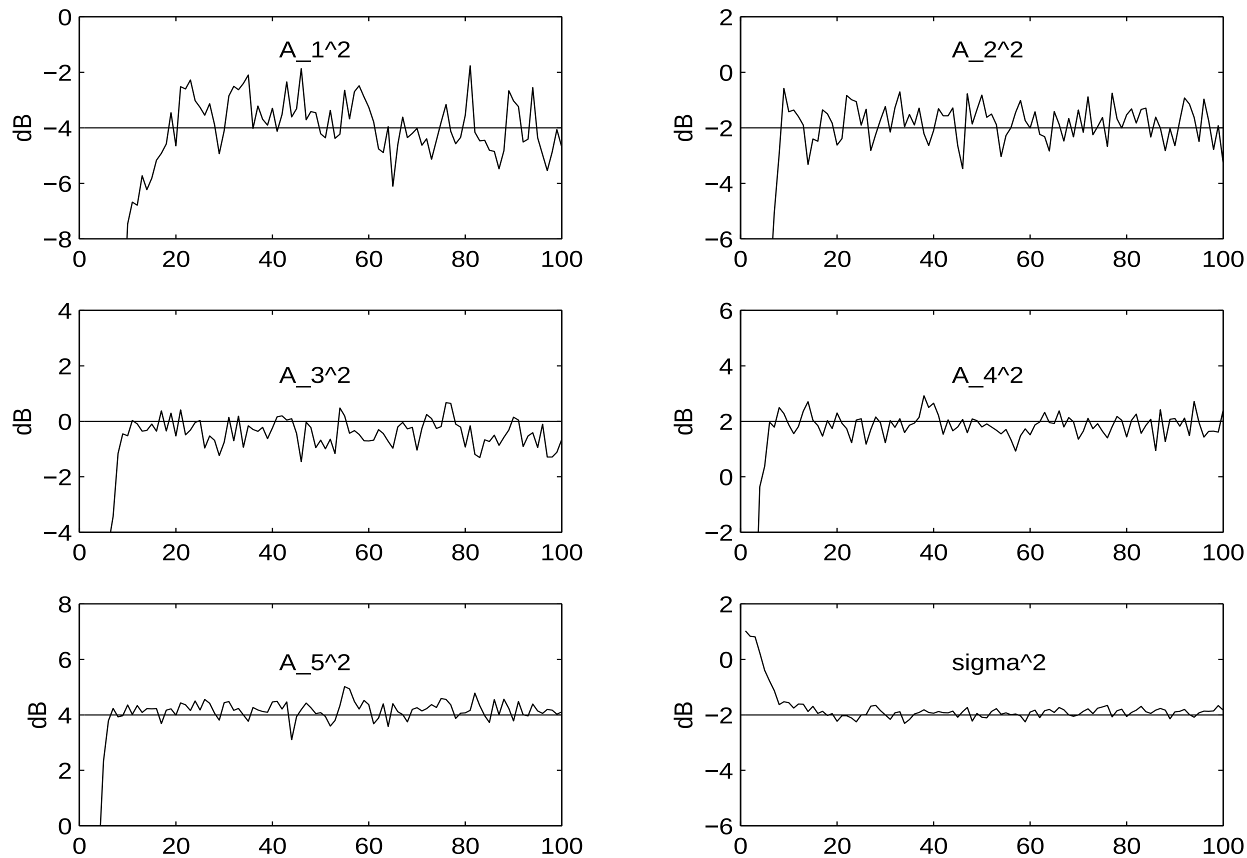

Figure 3: Samples drawn by the Gibbs multiuser detector in a Gaussian synchronous CDMA channel. $A_1^2 = -4\text{dB}$, $A_2^2 = -2\text{dB}$, $A_3^3 = 0\text{dB}$, $A_4^2 = 2\text{dB}$, $A_5^2 = 4\text{dB}$, and $\sigma^2 = -2\text{dB}$.

values of these quantities are also shown in the same figure as the straight lines. It is seen that the Gibbs sampler converges fairly rapidly (within about 20 iterations).

Performance of the Adaptive Turbo Multiuser Detectors

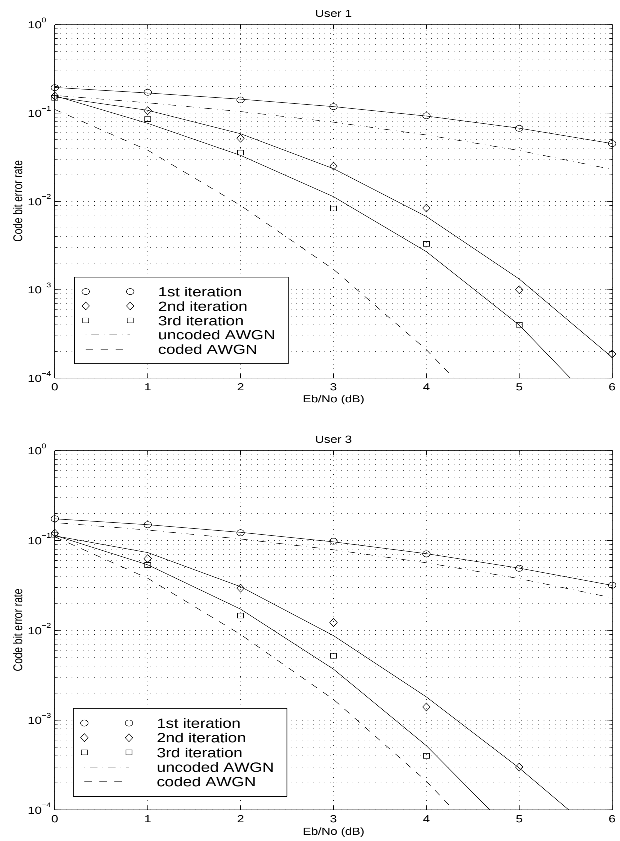

Figure 4: Bit error rate performance of the adaptive Turbo multiuser detector in a synchronous CDMA system with Gaussian noise. All users have the same amplitudes.

We now illustrate the performance of the adaptive Turbo multiuser detectors discussed in Section 6. The channel code for each user is a rate $\frac{1}{2}$ constraint length-5 convolutional code (with generators 23, 35 in octal notation). The interleaver of each user is indepedently 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 amplitudes. In computing the symbol probabilities, the Gibbs sampler is iterated 100 runs for each data block, with the first 50 iterations as the "burn-in" period. The symbol posterior probabilities are computed according to (24) with $n_0 = N = 50$.

Figure 4 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 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 signle-user performance in an AWGN channel within a few iterations.

Performance of the Code-constrained Gibbs Multiuser Detectors

Finally we consider the performance of the code-constrained Gibbs multiuser detectors discussed in Section 7. We assume that each user employs the (7,4) cyclic block code with eight possible codewords [32]:

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

The bit error rate performance of the code-constrained Gibbs multiuser detector in Gaussian noise is shown in Figure 6. In this case the Gibbs sampler draws a code word from $\mathcal{X}$ at each time, according to (33). 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.

8 Conclusions

In this paper, we have developed a new adaptive multiuser detection scheme which 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, a Markov Chain Monte Carlo procedure for computing Bayesian estimates. We have derived the adaptive multiuser detection algorithms for the Gaussian synchronous CDMA channel. The proposed

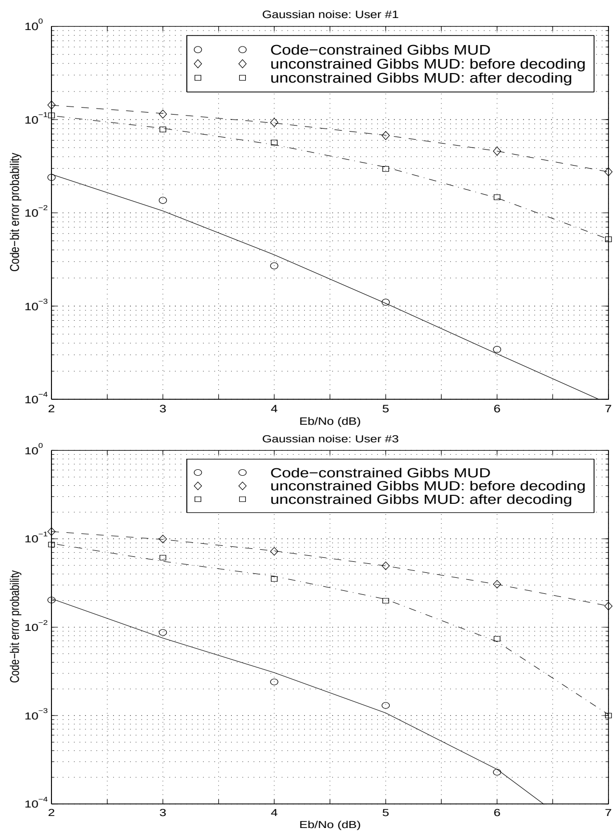

Figure 5: Bit error rate performance of the code-constrained Gibbs multiuser detector in a synchronous CDMA system with Gaussian noise. All users have the same amplitudes.

adaptive Bayesian multiuser detectors can incorporate the *a priori* symbol probabilities, and they produces 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. Future extensions to this work include generalizations of the techniques proposed here to asynchronous CDMA systems and to systems with multipath fading effects.

References

[1] C. Berrou and A. Glavieux. Near optimum error-correcting coding and decoding: Turbo codes. *IEEE Trans. Commun.*, COM-44(10), Oct. 1996.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correction coding and decoding: Turbo codes. In *Proc. 1993 International Conference on Communications (ICC'93)*, pages 1064–1070, 1993.

[3] K.S. Chan. Asymptotic behavior of the Gibbs sampler. *J. Amer. Stat. Assoc.*, 88:320–326, 1993.

[4] R. Chen and T.H. Li. Blind restoration of linearly degraded discrete signals by Gibbs sampling. *IEEE Trans. Sig. Proc.*, 43(10):2410–2413, Oct. 1995.

[5] R. Chen and J.S. Liu. Predictive updating methods with applications to Bayesian classification. *J. Royal Statist. Soc. B*, 58:397–415, 1995.

[6] U. Fawer and B. Aazhang. A multiuser receiver for code-division multiple-access communications over multipath channels. *IEEE Trans. Commun.*, COM-43(2/3/4):1556–1565, Feb./Mar./Apr. 1995.

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

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6(11):721–741, Nov. 1984.

[9] T.R. Giallorenzi and S.G. Wilson. Multiuser ML sequence estimator for convolutional coded asynchronous DS-CDMA systems. *IEEE Trans. Commun.*, COM-44(8):997–1008, Aug. 1996.

[10] T.R. Giallorenzi and S.G. Wilson. Suboptimum multiuser receivers for convolutionally coded asynchrnous DS-CDMA systems. *IEEE Trans. Commun.*, COM-44(9):1183–1196, Sep. 1996.

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

[12] J. Hagenauer. The Turbo principle: Tutorial introduction and state of the art. In *Proc. International Symposium on Turbo Codes and Related Topics*, pages 1–11, Brest, France, Sep. 1997.

[13] R.A. Iltis. An EKF-based joint estimator for interference, multipath and code delay in a DS spread-spectrum receiver. *IEEE Trans. Commun.*, 42(2/3/4):1288–1299, Feb./Mar./Apr. 1994.

[14] R.A. Iltis and L. Mailaender. An adaptive multiuser detector with joint amplitude and delay estimation. *IEEE J. Select. Areas Commun.*, 12(5):774–785, June 1994.

[15] 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*, 81:27–40, 1994.

[16] S.Y. Miller. *Detection and estimation in multiple-access channels*. PhD thesis, Department of Electrical Engineering, Princeton University, 1989.

[17] M. Moher. An iterative multiuser decoder for near-capacity communications. *IEEE Trans. Commun.*, COM-46(7):870–880, July 1998.

[18] T.K. Moon, Z. Xie, C.K. Rushforth, and R.T. Short. Parameter estimation in a multiuser communication system. *IEEE Trans. Commun.*, 42(8):2553–2560, Aug. 1994.

[19] H.V. Poor. On parameter estimation in DS/CDMA formats. In W.A. Porter and S.C. Kak, editors, *Advances in Communications and Signal Processing*. Springer-Verlag, New York, 1989.

[20] 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.*, 46(12):1693–1699, Dec. 1998.

[21] M.J. Schervish and B.P. Carlin. On the convergence of successive substitution sampling. *J. Cmp. Gr. St.*, 1:111–127, 1992.

[22] Y. Steinberg and H.V. Poor. Sequential amplitude estimation in multiuser communications. *IEEE Trans. Inform. Theory*, IT-40(1):11–20, Jan. 1994.

[23] E. Strom and S. Miller. Asynchronous DS-CDMA systems: Low complexity near-far resistant receiver and parameter estimation. Technical Reports, Department of Electrical Engineering, University of Florida, Jan. 1994.

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

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

[26] S. Verdú. *Optimal multiuser signal detection*. PhD thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1984.

[27] S. Verdú. Minimum probability of error for asynchronous Gaussian multiple-access channels. *IEEE Trans. Inform. Theory*, IT-32(1):85–96, Jan. 1986.

[28] S. Verdú. Computational complexity of optimum multiuser detection. *Algorithmica*, 4:303–312, 1989.

[29] S. Verdú. *Multiuser Detection*. Cambridge University Press: Cambridge, UK, 1998.

[30] J. von Neumann. Various techniques used in connection with random digit. *Natl. Bureau of Standards Appl. Math. Ser.*, 12:36–38, 1951.

[31] X. Wang and H.V. Poor. Iterative (Turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Trans. Commun.*, 47(7), July. 1999.

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

[33] W.H. Wong and F. Liang. Dynamic importance weighting in Monte Carlo and optimization. *Proc. Natl. Acad. Sci.*, 94:14220–14224, 1997.

[34] Z. Xie, C.K. Rushforth, R.T. Short, and T.K. Moon. Joint signal detection and parameter estimation in multiuser communications. *IEEE Trans. Commun.*, 41(8):1208–1216, Aug. 1993.