Blind Turbo Equalization in Gaussian and Impulsive Noise

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

***Abstract*—We consider the problem of simultaneous parameter estimation and restoration of finite-alphabet symbols that are blurred by an unknown linear intersymbol interference (ISI) channel and contaminated by additive Gaussian or non-Gaussian white noise with unknown parameters. Non-Gaussian noise is found in many wireless channels due to impulsive phenomena of radio-frequency interference. Bayesian inference of all unknown quantities is made from the blurred and noisy observations. 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. Blind Bayesian equalizers based on the Gibbs sampler are derived for both Gaussian ISI channel and impulsive ISI channel. A salient feature of the proposed blind Bayesian equalizers 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 blind Bayesian equalizer to refine its processing based on the information from the decoding stage and vice versa—a receiver structure termed as *blind turbo equalizer*.**

***Index Terms*—Bayesian inference, blind equalization, Gibbs sampler, impulsive noise, iterative processing.**

I. Introduction

In a band-limited digital communication system, the transmitted digital symbols are distorted by the base-band equivalent discrete-time linear finite impulse response (FIR) channel, causing intersymbol interference (ISI). Blind equalization refers to the reconstruction of transmitted symbols based on the noise-corrupted channel output without knowing the underlying FIR channel. The traditional approach to blind equalization in digital communications is to use a linear equalizer, i.e., an FIR transversal filter. Two families of methodologies for blind adaptation of a linear equalizer are the high-order statistics (HOS)-based techniques, (such as the Bussgang methods and the polyspectral methods), and the second-order statistics (SOS)-based techniques, (such as the cyclostationarity-based methods and the subspace-based methods). Discussions on these methodologies can be found in the articles collected in [15] and the more recent surveys [16], [43].

Since a linear equalizer may perform poorly in a severe ISI channel, some recent work has addressed nonlinear blind equalization techniques. For example, the algorithms in [21], [25], [26], [40] employ a sequence estimator and a bank of channel estimators and alternatively optimize with respect to data and channel. Sequence estimation is performed by a blind search of a modified trellis and channel estimation is accomplished by conditioning on survivor sequences in the trellis and constructing the corresponding maximum likelihood or minimum mean-square error channel estimate. Similar iterative methods for joint sequence detection and channel estimation are also found in [17], [27] and it is shown in [27] that these procedures are based on the expectation-maximization (EM) algorithm. Symbol-by-symbol maximum *a posteriori* probability (MAP)-based blind equalization schemes have also been considered. For instance, Bayesian blind equalization techniques are proposed in [22], [26], which combine recursive channel estimation with the MAP and the Bayesian decision feedback equalization methods for known channels introduced in [1]. Another Bayesian blind equalization method is proposed in [30], which is a Kalman-filter-like updating algorithm for joint symbol and channel estimation.

To date most of the work on equalization and blind equalization assumes that the channel ambient noise is Gaussian. However, in many physical channels such as urban and indoor radio channels [5], [6], [34], [35], [37] and underwater acoustic channels [8], [36], 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. (For recent measurement results of impulsive noise in outdoor/indoor mobile and portable radio communications, see [5], [6] and the references therein.) It is widely known that linear signal processing procedures are ineffective in combating non-Gaussian noise [28]. Hence, nonlinear equalizers are necessary for impulsive ISI channels. A few recent works address *training-based* nonlinear equalization techniques in impulsive channels. For example, in [9], [29] nonlinear equalizers based respectively on the radial basis function network and the EM algorithm are proposed.

Recently, iterative ("turbo") processing techniques have received considerable attention following the discovery of the powerful turbo codes [3], [4]. The so called turbo-principle can be successfully applied to many detection/decoding problems such as serial concatenated decoding, equalization, coded modulation, multi-user detection and joint source and channel decoding [24]. Turbo equalization was first proposed in [14], where it is assumed that the channel coefficients are known. In [45], a low-complexity turbo multi-user receiver is developed

Manuscript received November 22, 1999; revised January 23, 2001. This work was supported in part by the Interdisciplinary Research Initiatives Program, Texas A&M University. X. Wang was supported in part by NSF Grant CAREER CCR-9875314. R. Chen was supported in part by the U.S. National Science Foundation under Grant DMS 9626113.

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 Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL 60607 USA.

Publisher Item Identifier S 0018-9545(01)04892-7.

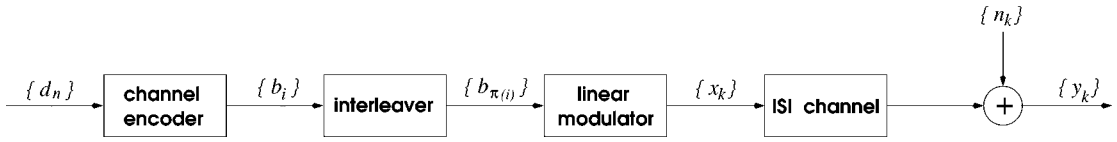

Fig. 1. A channel-coded communication system signaling through an intersymbol interference (ISI) channel with additive ambient noise.

for joint equalization and multi-user detection in CDMA systems, where it is also assumed that the user channels are known to the receiver.

In this paper, we present novel Bayesian blind equalization techniques for both Gaussian and impulsive ISI channels. The abovementioned previous "Bayesian" equalization schemes, e.g., [22], [26], [30], all make various approximating assumptions in deriving the equalizers and, therefore, they are not *true* Bayesian procedures. We consider Bayesian inference of all unknown quantities (e.g., channel states, symbol values, noise parameters) from the ISI-corrupted and noisy observations. A Markov chain Monte Carlo procedure called the Gibbs sampler is employed to calculate the Bayesian estimates. The performance of the proposed blind Bayesian equalizers is demonstrated via simulations. 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 blind equalization and decoding—so-called *blind turbo equalization*.

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 blind Bayesian equalization of Gaussian and impulsive ISI channels are treated in Sections IV and V respectively. In Section VI, a blind turbo equalization scheme is presented. Some discussions are found in Section VII. Simulation results are provided in Section VIII. Finally, Section IX contains the conclusions.

II. System Description

We consider a channel-coded communication system signaling through an intersymbol interference (ISI) channel with additive ambient noise. The block diagram of the transmitter-end of such a system is shown in Fig. 1. The binary information bits $\{d_n\}$ are encoded using some channel code (e.g., block code, convolutional code, or turbo code), resulting in a code bit stream $\{b_i\}$. 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 $\{b_{\pi(i)}\}$ are then passed to a linear modulator, where L code bits are mapped to one symbol. (e.g., PSK modulation or QAM modulation), yielding complex data symbols $\{x_k\}$. Each data symbol is then transmitted through an ISI channel. Suppose that a block of M symbols are transmitted. The discrete-time input–output relationship of the ISI channel is represented by the following linear finite impulse response (FIR) model

$$y_k = \sum_{i=0}^{Q-1} h_i^* x_{k-i} + n_k = \boldsymbol{h}^H \boldsymbol{x}_k + n_k, \qquad k = 0, \ldots, M-1. \tag{1}$$

In (1), y_k, x_k and n_k are, respectively, the received signal, the transmitted symbol, and ambient noise sample at time k; Q is the length of the channel memory; M is the size of the transmitted symbol block; $\{h_i^*\}_{i=0}^{Q-1}$ are the complex coefficients of the ISI channel; ^{H} denotes the conjugate transpose operation; and $\boldsymbol{h} \triangleq [h_0, h_1, \ldots, h_{Q-1}]^T$, $\boldsymbol{x}_k \triangleq [x_k, x_{k-1}, \ldots, x_{k-Q+1}]^T$.

It is assumed that the complex symbols $\{x_k\}$ are independent and they are drawn from a finite alphabet set $\mathcal{A} = \{a_1, a_2, \ldots, a_{|\mathcal{A}|}\}$. Define the following *a priori* probabilities of symbol x_k

$$\rho_{kj} \triangleq P(x_k = a_j), \qquad j = 1, \ldots, |\mathcal{A}|;\ k = 0, \ldots, M-1. \tag{2}$$

Note that when no prior information is available, then $\rho_{kj} = 1/|\mathcal{A}|$, i.e., all symbols are equally likely.

It is further assumed that the additive ambient channel noise $\{n_k\}$ is a sequence of zero-mean independent and identically distributed (i.i.d.) complex random variables, and it is independent of the symbol sequence $\{x_k\}$. 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 sample n_k is assumed to have a complex Gaussian distribution, i.e.,

$$n_k \sim \mathcal{N}_c(0, \sigma^2) \tag{3}$$

where σ^2 is the noise variance. For the latter case, the noise sample n_k is assumed to have a two-term Gaussian mixture distribution, i.e.,

$$n_k \sim (1-\epsilon)\mathcal{N}_c(0, \sigma_1^2) + \epsilon \mathcal{N}_c(0, \sigma_2^2) \tag{4}$$

with $0 < \epsilon < 1$ and $\sigma_2^2 > \sigma_1^2$. Here, the term $\mathcal{N}_c(0, \sigma_1^2)$ represents the nominal ambient noise and the term $\mathcal{N}_c(0, \sigma_2^2)$ represents an impulsive component, with ϵ representing the probability that impulses occur. The total noise variance under distribution (4) is given by

$$\sigma^2 = (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2. \tag{5}$$

Denote $\boldsymbol{Y} \triangleq \{y_0, y_1, \ldots, y_{M-1}\}$. In Sections IV and V, we consider the problem of estimating the *a posteriori* probabilities of the transmitted symbols

$$P(x_k = a_j | \boldsymbol{Y}), \qquad j = 1, \ldots, |\mathcal{A}|;\ k = 0, \ldots, M-1 \tag{6}$$

based on the received signals $\boldsymbol{Y}$ and the prior information $\{\rho_{kj}\}_{j=1;\,k=0}^{|\mathcal{A}|;\,M-1}$, without knowing the channel $\boldsymbol{h}$ 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 $\{d_n\}$ shown in Fig. 1, which will be discussed in Section VI.

III. The Gibbs Sampler

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., $p(\theta_j|\boldsymbol{Y})$, $1 \leq j \leq d$. Direct evaluation involves integrating out the rest of the parameters from the joint *a posteriori* density, i.e.,

$$p(\theta_j|\boldsymbol{Y}) = \iint \cdots \int p(\boldsymbol{\theta}|\boldsymbol{Y})\, d\theta_1 \cdots d\theta_{j-1}\, d\theta_{j+1} \cdots d\theta_d. \tag{7}$$

In most cases, such a direct evaluation is computational infeasible, especially when the parameter dimension d is large. The Gibbs sampler [18] 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 $p(\boldsymbol{\theta}|\boldsymbol{Y})$ 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 $p(\theta_1|\theta_2^{(n)}, \ldots, \theta_d^{(n)}, \boldsymbol{Y})$.
- Draw sample $\theta_2^{(n+1)}$ from $p(\theta_2|\theta_1^{(n+1)}, \theta_3^{(n)}, \ldots, \theta_d^{(n)}, \boldsymbol{Y})$

$\vdots$

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

The convergence behavior of the Gibbs sampler is investigated in [10], [18], [20], [33], [39], and [42] and general conditions are given for the following two results:

- the distribution of $\boldsymbol{\theta}^{(n)}$ converges geometrically to $p(\boldsymbol{\theta}|\boldsymbol{Y})$, as $n \longrightarrow \infty$;
- $(1/N)\sum_{n=1}^{N} f(\boldsymbol{\theta}^{(n)}) \xrightarrow{\text{a.s.}} \int f(\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\boldsymbol{Y})\, d\boldsymbol{\theta}$, 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.

IV. Blind Bayesian Equalization in Gaussian Noise

In this section, we consider the problem of computing the *a posteriori* symbol probabilities in (6), under the assumption that the ambient noise distribution is complex Gaussian. That is, the probability density function (pdf) of n_k in (1) is given by

$$p(n_k) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{|n_k|^2}{\sigma^2}\right). \tag{8}$$

Denote $\boldsymbol{X} \triangleq \{x_0, x_1, \ldots, x_{M-1}\}$. The problem is solved under a Bayesian framework. First, the unknown quantities $\boldsymbol{h}$, σ^2 and $\boldsymbol{X}$ are regarded as realizations of random variables with some prior distributions. The Gibbs sampler, is then employed to calculate the maximum *a posteriori* (MAP) estimates of these unknowns.

A. Bayesian Inference

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

$$\begin{aligned} p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right) &\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma^2, \boldsymbol{X}\right)\, p(\boldsymbol{h})\, p\left(\sigma^2\right)\, p(\boldsymbol{X}) \\ &\propto \left(\frac{1}{\sigma^2}\right)^M \exp\left(-\frac{1}{\sigma^2}\sum_{k=0}^{M-1}\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2\right) \\ &\quad \cdot p(\boldsymbol{h}) p\left(\sigma^2\right) p(\boldsymbol{X}). \end{aligned} \tag{9}$$

The *a posteriori* probabilities (6) of the transmitted symbols can then be calculated from the joint posterior distribution (9) according to

$$\begin{aligned} P(x_k = a_j|\boldsymbol{Y}) &= \sum_{\boldsymbol{X}:\, x_k = a_j} p(\boldsymbol{X}|\boldsymbol{Y}) \\ &= \sum_{\boldsymbol{X}:\, x_k = a_j} \int p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right)\, d\boldsymbol{h}\, d\sigma^2. \end{aligned} \tag{10}$$

Clearly, the computation in (10) involves $|\mathcal{A}|^{M-1}$ multidimensional integrals, which is certainly infeasible for any practical implementations. To avoid the direct evaluation of the Bayesian estimate (10), we resort to the Gibbs sampler discussed in Section III. The basic idea is to generate ergodic random samples $\{\boldsymbol{h}^{(n)}, \sigma^{2(n)}, \boldsymbol{X}^{(n)}: n = 0, 1, \ldots\}$ from the posterior distribution (9), and then to average $\{x_k^{(n)}: n = 0, 1, \ldots\}$ to obtain an approximation of the *a posteriori* probabilities in (10).

B. Prior Distributions

1) General Considerations:

a) 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 play a minimal role in the posterior distribution. Such priors are termed as *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.

b) 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 follows the same parametric form 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. 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 [7], [19], [31]).

2) Prior Distributions of the Unknowns: Following the general guidelines in Bayesian analysis [7], [19], [31], we choose the conjugate prior distributions for the unknown parameters $p(\boldsymbol{h})$, $p(\sigma^2)$ and $p(\boldsymbol{X})$ as follows.

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

$$p(\boldsymbol{h}) \sim \mathcal{N}(\boldsymbol{h}_0, \boldsymbol{\Sigma}_0). \tag{11}$$

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

$$\begin{aligned} p\left(\sigma^2\right) &= \frac{(\nu_0\lambda_0)^{\nu_0}}{\Gamma(\nu_0)}\left(\frac{1}{\sigma^2}\right)^{\nu_0+1}\exp\left(-\frac{\nu_0\lambda_0}{\sigma^2}\right) \\ &\sim \chi^{-2}(2\nu_0, \lambda_0) \end{aligned} \tag{12}$$

or

$$\frac{2\nu_0\lambda_0}{\sigma^2} \sim \chi^2(2\nu_0). \tag{13}$$

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

$$p(\boldsymbol{X}) = \prod_{k=0}^{M-1}\prod_{j=1}^{|\mathcal{A}|}\rho_{kj}^{\delta_{kj}} \tag{14}$$

where δ_{kj} is the indicator such that $\delta_{kj} = 1$ if $x_k = a_j$ and $\delta_{kj} = 0$ if $x_k \neq a_j$.

C. Conditional Posterior Distributions

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

1) The conditional distribution of the channel response $\boldsymbol{h}$ given σ^2, $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$p\left(\boldsymbol{h}|\sigma^2, \boldsymbol{X}, \boldsymbol{Y}\right) \sim \mathcal{N}_c(\boldsymbol{h}_*, \boldsymbol{\Sigma}_*) \tag{15}$$

with

$$\boldsymbol{\Sigma}_*^{-1} \triangleq \boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\boldsymbol{x}_k\boldsymbol{x}_k^H \tag{16}$$

and

$$\boldsymbol{h}_* \triangleq \boldsymbol{\Sigma}_*\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{h}_0 + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\boldsymbol{x}_k y_k^*\right). \tag{17}$$

(Recall that $\boldsymbol{x}_k \triangleq [x_k, \ldots, x_{k-Q+1}]^T$.)

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

$$p\left(\sigma^2|\boldsymbol{h}, \boldsymbol{X}, \boldsymbol{Y}\right) \sim \chi^{-2}\left(2[\nu_0 + M], \frac{\nu_0\lambda_0 + s^2}{\nu_0 + M}\right) \tag{18}$$

with

$$s^2 \triangleq \sum_{k=0}^{M-1}\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2. \tag{19}$$

3) The conditional symbol probabilities given $\boldsymbol{h}$, σ^2, $\boldsymbol{X}_{[-k]}$ and $\boldsymbol{Y}$ can be obtained from [where $\boldsymbol{X}_{[-k]}$ denotes the set $\{x_0, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{M-1}\}$]

$$\begin{aligned} &\frac{P\left(x_k = a_j|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}{P\left(x_k = a_i|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)} \\ &\quad= \frac{\rho_{kj}}{\rho_{ki}}\cdot\exp\left\{-\frac{1}{\sigma^2}\sum_{l=k}^{\overline{k}}\left[|h_{l-k}|^2\left(|a_j|^2 - |a_i|^2\right)\right.\right. \\ &\qquad\left.\left.-2\Re\left\{h_{l-k}\left(y_l - \sum_{\substack{m=0\\ m\neq l-k}}^{Q-1}h_m^* x_{l-m}\right)(a_j^* - a_i^*)\right\}\right]\right\} \\ &\qquad k = 0, \ldots, M-1; \quad i, j = 1, \ldots, |\mathcal{A}| \end{aligned} \tag{20}$$

where $\overline{k} = \min\{k + Q - 1, M - 1\}$.

Note that in the case that the initial symbols $x_{-(Q-1)}, \ldots, x_{-1}$ are unknown, they can be included in the analysis. For $k = -(Q-1), \ldots, -1$, we have

$$\begin{aligned} &\frac{P\left(x_k = a_j|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}{P\left(x_k = a_i|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)} \\ &\quad= \frac{\rho_{kj}}{\rho_{ki}}\cdot\exp\left\{-\frac{1}{\sigma^2}\sum_{l=0}^{k+Q-1}\left[|h_{l-k}|^2\left(|a_j|^2 - |a_i|^2\right)\right.\right. \\ &\qquad\left.\left.-2\Re\left\{h_{l-k}\left(y_l - \sum_{\substack{m=0\\ m\neq l-k}}^{Q-1}h_m^* x_{l-m}\right)(a_j^* - a_i^*)\right\}\right]\right\} \\ &\qquad i, j = 1, \ldots, |\mathcal{A}|. \end{aligned} \tag{21}$$

D. Gibbs Blind Equalizer in Gaussian Noise

Using the above conditional posterior distributions, the Gibbs sampling implementation of the blind Bayesian equalizer in Gaussian noise proceeds iteratively as follows. Given the initial values of the unknown quantities $\{\boldsymbol{h}^{(0)}, \sigma^{2(0)}, \boldsymbol{X}^{(0)}\}$ drawn from their prior distributions and for $n = 1, 2, \ldots$

1) draw $\boldsymbol{h}^{(n)}$ from $p(\boldsymbol{h}|\sigma^{2(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (15);
2) draw $\sigma^{2(n)}$ from $p(\sigma^2|\boldsymbol{h}^{(n)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (18);
3) for $k = 0, 1, \ldots, M-1$, draw $x_k^{(n)}$ from $P(x_k|\boldsymbol{h}^{(n)}, \sigma^{2(n)}, \boldsymbol{X}_{[-k]}^{(n)}, \boldsymbol{Y})$ given by (20), where

$$\boldsymbol{X}_{[-k]}^{(n)} \triangleq \left[x_0^{(n)}, \ldots, x_{k-1}^{(n)}, x_{k+1}^{(n-1)}, \ldots, x_{M-1}^{(n-1)}\right].$$

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 marginal *a posteriori* symbol probabilities in (10) are approximated as

$$P(x_k = a_j|\boldsymbol{Y}) \cong \frac{1}{N}\sum_{n=n_0+1}^{n_0+N}\delta_{kj}^{(n)} \tag{22}$$

where $\delta_{kj}^{(n)}$ is the indicator such that $\delta_{kj}^{(n)} = 1$ if $x_k^{(n)} = a_j$ and $\delta_{kj}^{(n)} = 0$ if $x_k^{(n)} \neq a_j$. Furthermore, if desired, the estimates of the channel response $\boldsymbol{h}$ and the noise variance σ^2 can also be obtained from the corresponding sample means

$$E\{\boldsymbol{h}|\boldsymbol{Y}\} \cong \frac{1}{N} \sum_{n=n_0+1}^{n_0+N} \boldsymbol{h}^{(n)} \tag{23}$$

and

$$E\left\{\sigma^2|\boldsymbol{Y}\right\} \cong \frac{1}{N} \sum_{n=n_0+1}^{n_0+N} \sigma^{2(n)}. \tag{24}$$

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 blind equalizer is $O((Q^2+M)N)$, where N is the number of iterations in the Gibbs sampling (in our simulations $N = 100$). That is, it has a term which is quadratic with respect to the channel memory size Q [due to the inversion of the positive definite symmetric matrix in (16)], and a term which is *linear* with respect to the symbol block size M. This is a considerable reduction in computational complexity compared with the direct implementation of the Bayesian symbol estimate (10)], which is on the order of $O(|\mathcal{A}|^M)$.

V. Blind Bayesian Equalization in Impulsive Noise

So far we have assumed that the distribution of the ambient channel noise is Gaussian, as did in most of the previous work on equalization. However, in many realistic communication channels, especially wireless channels, the ambient noise is known to be decidedly non-Gaussian, due to impulsive phenomena [5], [6]. In this section, we develop the Gibbs blind equalizer in impulsive noise. It is assumed that the noise samples $\{n_k\}_{k=0}^{M-1}$ in (1) are independent with the common two-term Gaussian mixture pdf, given by

$$p(n_k) = \frac{1-\epsilon}{\pi\sigma_1^2} \exp\left(-\frac{|n_k|^2}{\sigma_1^2}\right) + \frac{\epsilon}{\pi\sigma_2^2} \exp\left(-\frac{|n_k|^2}{\sigma_2^2}\right) \tag{25}$$

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 [36], [47] 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 I_k, $k = 0, \ldots, M-1$,

$$I_k = \begin{cases} 1, & \text{if } n_k \sim \mathcal{N}_c\left(0, \sigma_1^2\right) \\ 2, & \text{if } n_k \sim \mathcal{N}_c\left(0, \sigma_2^2\right). \end{cases} \tag{26}$$

Denote $\boldsymbol{I} \triangleq \{I_0, I_1, \ldots, I_{M-1}\}$. The unknown quantities in this case are $(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X})$. The joint posterior distribution of these unknown quantities based on the received signal $\boldsymbol{Y}$ takes the form of

$$\begin{aligned} &p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right) \\ &\quad \propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\right) \\ &\qquad \cdot p(\boldsymbol{h}) p\left(\sigma_1^2\right) p\left(\sigma_2^2\right) p(\epsilon) p(\boldsymbol{I}|\epsilon) p(\boldsymbol{X}) \\ &\quad \propto \left(\frac{1}{\sigma_1^2}\right)^{n_1} \left(\frac{1}{\sigma_2^2}\right)^{n_2} \exp\left(-\sum_{k=0}^{M-1} \frac{\left|y_k - \boldsymbol{h}^H \boldsymbol{x}_k\right|^2}{\sigma_{I_k}^2}\right) \\ &\qquad \cdot p(\boldsymbol{h}) p\left(\sigma_1^2\right) p\left(\sigma_2^2\right) p(\epsilon) p(\boldsymbol{I}|\epsilon) p(\boldsymbol{X}) \end{aligned} \tag{27}$$

where n_1 is number of ones in $\boldsymbol{I}$ and $n_2 = M - n_1$. We next specify the prior distributions of the unknown quantities in (27).

As in the case of Gaussian noise, the prior distributions $p(\boldsymbol{h})$ and $p(\boldsymbol{X})$ are given respectively by (11) and (14). For the noise variances σ_i^2, $i = 1, 2$, independent inverse chi-square distributions are assumed, i.e.,

$$p(\sigma_i^2) \sim \chi^{-2}(2\nu_i, \lambda_i), \qquad i = 1, 2, \quad \text{with } \nu_1\lambda_1 < \nu_2\lambda_2. \tag{28}$$

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

$$p(\epsilon) = \frac{\Gamma(a_0+b_0)}{\Gamma(a_0)\Gamma(b_0)} \epsilon^{a_0-1}(1-\epsilon)^{b_0-1} \sim \text{Beta}(a_0, b_0). \tag{29}$$

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_k is then

$$P(I_k = 1|\epsilon) = 1-\epsilon, \quad \text{and} \quad P(I_k = 2|\epsilon) = \epsilon \tag{30}$$

$$\Rightarrow \quad P(\boldsymbol{I}|\epsilon) = (1-\epsilon)^{n_1}\epsilon^{n_2}. \tag{31}$$

B. Conditional Posterior Distributions

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

1) The conditional distribution of the channel response $\boldsymbol{h}$ given σ_1^2, σ_2^2, ϵ, $\boldsymbol{I}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$p\left(\boldsymbol{h}|\sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right) \sim \mathcal{N}_c(\boldsymbol{h}_*, \boldsymbol{\Sigma}_*) \tag{32}$$

with

$$\boldsymbol{\Sigma}_*^{-1} \triangleq \boldsymbol{\Sigma}_0^{-1} + \sum_{k=0}^{M-1} \frac{1}{\sigma_{I_k}^2} \boldsymbol{x}_k \boldsymbol{x}_k^H \tag{33}$$

and

$$\boldsymbol{h}_* \triangleq \boldsymbol{\Sigma}_* \left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{h}_0 + \sum_{k=0}^{M-1} \frac{1}{\sigma_{I_k}^2} \boldsymbol{x}_k y_k^*\right). \tag{34}$$

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

$$p\left(\sigma_i^2|\boldsymbol{h}, \sigma_{\bar{i}}^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right) \sim \chi^{-2}\left(2[\nu_i + n_i], \frac{\nu_i\lambda_i + s_i^2}{\nu_i + n_i}\right) \quad (35)$$

with

$$s_i^2 \triangleq \sum_{k=0}^{M-1} \left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2 1_{\{I_k=i\}}, \qquad i = 1, 2 \quad (36)$$

where

n_i the number of is in $\boldsymbol{I}$;
i 1, 2, $n_1 + n_2 = M$;
$1_{\{I_k=i\}}$ is the indicator function such that $1_{\{I_k=i\}} = 1$ if $I_k = i$ and $1_{\{I_k=i\}} = 0$ if $I_k \neq i$.

3) The conditional symbol probabilities given $\boldsymbol{h}$, σ_1^2, σ_2^2, ϵ, $\boldsymbol{I}$, $\boldsymbol{X}_{[-k]}$ and $\boldsymbol{Y}$ can be obtained from [where $\boldsymbol{X}_{[-k]}$ denotes the set $\{x_0, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{M-1}\}$]

$$\frac{P\left(x_k = a_j|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}{P\left(x_k = a_i|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)} = \frac{\rho_{kj}}{\rho_{ki}} \cdot \exp\left\{-\sum_{l=k}^{\bar{k}} \frac{1}{\sigma_{I_l}^2}\left[|h_{l-k}|^2\left(|a_j|^2 - |a_i|^2\right) - 2\Re\left\{h_{l-k}\left(y_l - \sum_{\substack{m=0 \\ m\neq l-k}}^{Q-1} h_m^* x_{l-m}\right)(a_j^* - a_i^*)\right\}\right]\right\}$$
$$k = 0, \ldots, M-1; \;\; i, j = 1, \ldots, |\mathcal{A}| \quad (37)$$

where $\bar{k} = \min\{k + Q - 1, M - 1\}$.

4) The conditional distribution of the indicator I_k, given $\boldsymbol{h}$, σ_1^2, σ_2^2, ϵ, $\boldsymbol{I}_{[-k]}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$ is obtained from [where $\boldsymbol{I}_{[-k]}$ denotes the set $\{I_0, \ldots, I_{k-1}, I_{k+1}, \ldots, I_{M-1}\}$]

$$\frac{P\left(I_k = 1|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{[-k]}, \boldsymbol{X}, \boldsymbol{Y}\right)}{P\left(I_k = 2|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{[-k]}, \boldsymbol{X}, \boldsymbol{Y}\right)} = \frac{1-\epsilon}{\epsilon} \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \exp\left[\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right]. \quad (38)$$

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

$$p\left(\epsilon|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right) = \text{Beta}(a_0 + n_2, b_0 + n_1). \quad (39)$$

C. Gibbs Blind Equalizer in Impulsive Noise

Using the above conditional posterior distributions, the Gibbs sampling implementation of the blind Bayesian equalizer in impulsive noise proceeds iteratively as follows. Given initial values of the unknown quantities $\{\boldsymbol{h}^{(0)}, \sigma_1^{2(0)}, \sigma_2^{2(0)}, \epsilon^{(0)}, \boldsymbol{I}^{(0)}, \boldsymbol{X}^{(0)}\}$ drawn from their prior distributions, and for $n = 1, 2, \ldots$

1) Draw $\boldsymbol{h}^{(n)}$ from $p(\boldsymbol{h}|\sigma_1^{2(n-1)}, \sigma_2^{2(n-1)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (32).
2) Draw $\sigma_1^{2(n)}$ from $p(\sigma_1^2|\boldsymbol{h}^{(n)}, \sigma_2^{2(n-1)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (36); draw $\sigma_2^{2(n)}$ from $p(\sigma_2^2|\boldsymbol{h}^{(n)}, \sigma_1^{2(n)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ given by (36).
3) For $k = 0, 1, \ldots, M-1$, draw $x_k^{(n)}$ from $P(x_k|\boldsymbol{h}^{(n)}, \sigma_1^{2(n)}, \sigma_2^{2(n)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}_{[-k]}^{(n)}, \boldsymbol{Y})$ given by (37), where

$$\boldsymbol{X}_{[-k]}^{(n)} = \left[x_1^{(n)}, \ldots, x_{k-1}^{(n)}, x_{k+1}^{(n-1)}, \ldots, x_{M-1}^{(n-1)}\right].$$

4) For $k = 0, 1, \ldots, M-1$, draw $I_k^{(n)}$ from $P(I_k|\boldsymbol{h}^{(n)}, \sigma_1^{2(n)}, \sigma_2^{2(n)}, \epsilon^{(n-1)}, \boldsymbol{I}_{[-k]}^{(n)}, \boldsymbol{X}^{(n)}, \boldsymbol{Y})$ given by (38), where

$$\boldsymbol{I}_{[-k]}^{(n)} = \left[I_1^{(n)}, \ldots, I_{k-1}^{(n)}, I_{k+1}^{(n-1)}, \ldots, I_{M-1}^{(n-1)}\right].$$

5) Draw $\epsilon^{(n)}$ from $p(\epsilon|\boldsymbol{h}^{(n)}, \sigma_1^{2(n)}, \sigma_2^{2(n)}, \boldsymbol{I}^{(n)}, \boldsymbol{X}^{(n)}, \boldsymbol{Y})$ given by (39).

As in the case of Gaussian noise, the *a posteriori* symbol probabilities $P(x_k = a_j|\boldsymbol{Y})$ are computed using (22). The *a posteriori* means and variances of the other unknown quantities can also be computed, similar to (23) and (24).

VI. Iterative Joint Blind Equalization and Decoding—Blind Turbo Equalization

In this section, we consider employing iterative equalization and decoding to improve the performance of the blind Bayesian equalizer. Because they utilize the *a priori* symbol probabilities, and they produce symbol (or bit) *a posteriori* probabilities, the blind Bayesian equalizers developed in this paper are well suited for iterative processing which allows the blind equalizer to refine its processing based on the information from the decoding stage and vice versa.

The iterative (turbo) receiver structure is shown in Fig. 2. It consists of two stages: the blind Bayesian equalizer developed in the previous sections, followed by a soft-input soft-output channel decoder. The two stages are separated by deinterleavers and interleavers. As discussed in the previous sections, the blind Bayesian equalizer delivers the *a posteriori* symbol probabilities $\{P(x_k = a_j|\boldsymbol{Y})\}_{j=1;k=0}^{|\mathcal{A}|;M-1}$. Based on these, we first compute the *a posteriori* log-likelihood ratios (LLRs) of a transmitted "+1" and a transmitted "−1" for the interleaved *code bits* $\{b_{\pi(i)}\}_{i=0}^{LM-1}$ [cf. Fig. 1]. (Recall that L code bits are mapped to one symbol.) Assume that the code bit $b_{\pi(i)}$ is mapped to symbol $x_{\kappa(\pi(i))}$, where $0 \leq \kappa(\pi(i)) < M$. Then the LLR of this code bit is given by

$$\Lambda_1\left(b_{\pi(i)}\right) \triangleq \log\frac{P\left(b_{\pi(i)} = +1|\boldsymbol{Y}\right)}{P\left(b_{\pi(i)} = -1|\boldsymbol{Y}\right)} = \log\frac{\sum_{a_j\in\mathcal{A}:\, x_{\kappa(\pi(i))}=a_j,\, b_{\pi(i)}=+1} P\left(x_{\kappa(\pi(i))} = a_j|\boldsymbol{Y}\right)}{\sum_{a_j\in\mathcal{A}:\, x_{\kappa(\pi(i))}=a_j,\, b_{\pi(i)}=-1} P\left(x_{\kappa(\pi(i))} = a_j|\boldsymbol{Y}\right)}. \quad (40)$$

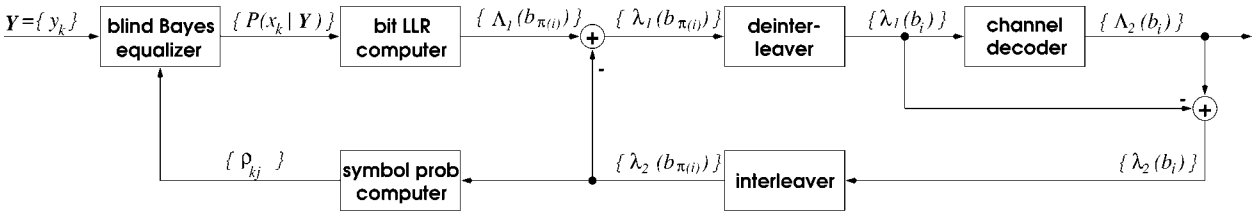

Fig. 2. Iterative processing for joint blind Bayesian equalization and decoding—blind turbo equalizer.

Using the Bayes' rule, (40) can be written as

$$\Lambda_1\left(b_{\pi(i)}\right) = \underbrace{\log \frac{p\left[\boldsymbol{Y}|b_{\pi(i)}=+1\right]}{p\left[\boldsymbol{Y}|b_{\pi(i)}=-1\right]}}_{\lambda_1(b_{\pi(i)})} + \underbrace{\log \frac{P\left(b_{\pi(i)}=+1\right)}{P\left(b_{\pi(i)}=-1\right)}}_{\lambda_2^p(b_{\pi(i)})} \tag{41}$$

where the second term in (41), denoted by $\lambda_2^p(b_{\pi(i)})$, represents the *a priori* LLR of the code bit $b_{\pi(i)}$, which is computed by the channel decoder in the previous iteration, interleaved and then fed back to the blind Bayesian equalizer. (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 $\lambda_2^p(b_{\pi(i)}) = 0$, for $0 \le i < LM$. The first term in (41), denoted by $\lambda_1(b_{\pi(i)})$, represents the *extrinsic* information delivered by the blind Bayesian equalizer based on the received signals $\boldsymbol{Y}$, the structure of the ISI-distorted signal given by (1), and the prior information about all other code bits. The extrinsic information $\lambda_1(b_{\pi(i)})$, which is not influenced by the *a priori* information $\lambda_2^p(b_{\pi(i)})$ 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 $\{\lambda_1^p(b_i)\}_{i=0}^{LM-1}$, and the structure of the channel code, the soft-input soft-output channel decoder computes the *a posteriori* LLR of each code bit

$$\begin{aligned}\Lambda_2(b_i) &\triangleq \log \frac{P\left(b_i=+1|\{\lambda_1^p(b_l)\}_{l=0}^{LM-1};\ \text{decoding}\right]}{P\left(b_i=-1|\{\lambda_1^p(b_l\}_{l=0}^{LM-1};\ \text{decoding}\right]} \\ &= \lambda_2(b_i) + \lambda_1^p(b_i), \qquad i=0,\ldots,LM-1. \end{aligned} \tag{42}$$

It is seen from (42) that the output of the soft-input soft-output channel decoder is the sum of the prior information $\lambda_1^p(b_i)$, and the *extrinsic* information $\lambda_2(b_i)$ delivered by the channel decoder. This extrinsic information is the information about the code bit b_i gleaned from the prior information about the other code bits, $\{\lambda_1^p(b_l)\}_{l\neq i}$ 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 $\{\lambda_2(b_i)\}_{i=0}^{LM-1}$ is then used to compute the *a priori* symbol distributions $\{\rho_{kj}\}$ defined in (6). Assume that a block of L bits $(\beta_{j,0}\beta_{j,1}\ldots\beta_{j,L-1})$, $\beta_{j,l}\in\{+1,-1\}$, is mapped to symbol a_j, for $j=1,\ldots,|\mathcal{A}|$. Denote $\iota(k,l)$ as the code bit index of the lth bit in the kth symbol, where $0\le k<M$, $0\le l<L$ and $0\le \iota(k,l)<LM$. The prior symbol probability is then given by

$$\rho_{kj} \triangleq P(x_k=a_j) = \prod_{l=0}^{L-1} P\left(b_{\iota(k,l)}=\beta_{j,l}\right). \tag{43}$$

The code bit probabilities in (43) can be computed from the corresponding LLRs as follows. Since $\lambda_2^p(b_i) = \log(P(b_i=+1)/P(b_i=-1))$, after some manipulations we have for $\beta\in\{+1,-1\}$

$$\begin{aligned} P(b_i=\beta) &= \frac{\exp\left[\beta\,\lambda_2^p(b_i)\right]}{1+\exp\left[\beta\,\lambda_2^p(b_i)\right]} \\ &= \frac{\exp\left[\frac{1}{2}\,\beta\,\lambda_2^p(b_i)\right]}{\exp\left[-\frac{1}{2}\,\beta\,\lambda_2^p(b_i)\right]+\exp\left[\frac{1}{2}\,\beta\,\lambda_2^p(b_i)\right]} \\ &= \frac{\cosh\left[\frac{1}{2}\,\lambda_2^p(b_i)\right]\left[1+\beta\,\tanh\left(\frac{1}{2}\,\lambda_2^p(b_i)\right)\right]}{2\cosh\left[\frac{1}{2}\,\lambda_2^p(b_i)\right]} \\ &= \tfrac{1}{2}\left[1+\beta\,\tanh\left(\frac{1}{2}\,\lambda_2^p(b_i)\right)\right]. \end{aligned} \tag{44}$$

The symbol probabilities $\{\rho_{kj}\}_{j=1;\,k=0}^{|\mathcal{A}|;\,M-1}$ are then fed back to the blind Bayesian equalizer as the prior information for the next iteration. Note that at the first iteration, the extrinsic information $\{\lambda_1(b_i)\}$ and $\{\lambda_2(b_i)\}$ 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.

VII. Discussions

1) Shift and Phase Ambiguities: Blind deconvolution problem, in general, can only be solved up to a time-delay ambiguity and sometimes also up to a phase ambiguity [2], [13], [32] if no further restrictions are imposed on the filter coefficients $\boldsymbol{h}$. In particular, when $h_i\approx 0$ for $i=q,\ldots,Q-1$ in (1), the time delay of the input signal $\{x_i\}$ is essentially unidentifiable. In fact, in this case, the models $y_k=\sum_{i=0}^{q-1} h_{i-\tau}^* x_{k+\tau-i}+n_k$, for $\tau=1,\ldots,q-1$ are all practically equivalent to (1). As a result, the posterior distribution can be an equally weighted mixture of several distributions, each corresponding to a particular time delay. In this case, estimators based on the marginal distribution cannot be used. The global MAP is a possible alternative, but it is difficult to obtain in high-dimensional cases. Furthermore, if the symbol alphabet is symmetric about zero, the blind equalizer is also subject to a phase ambiguity.

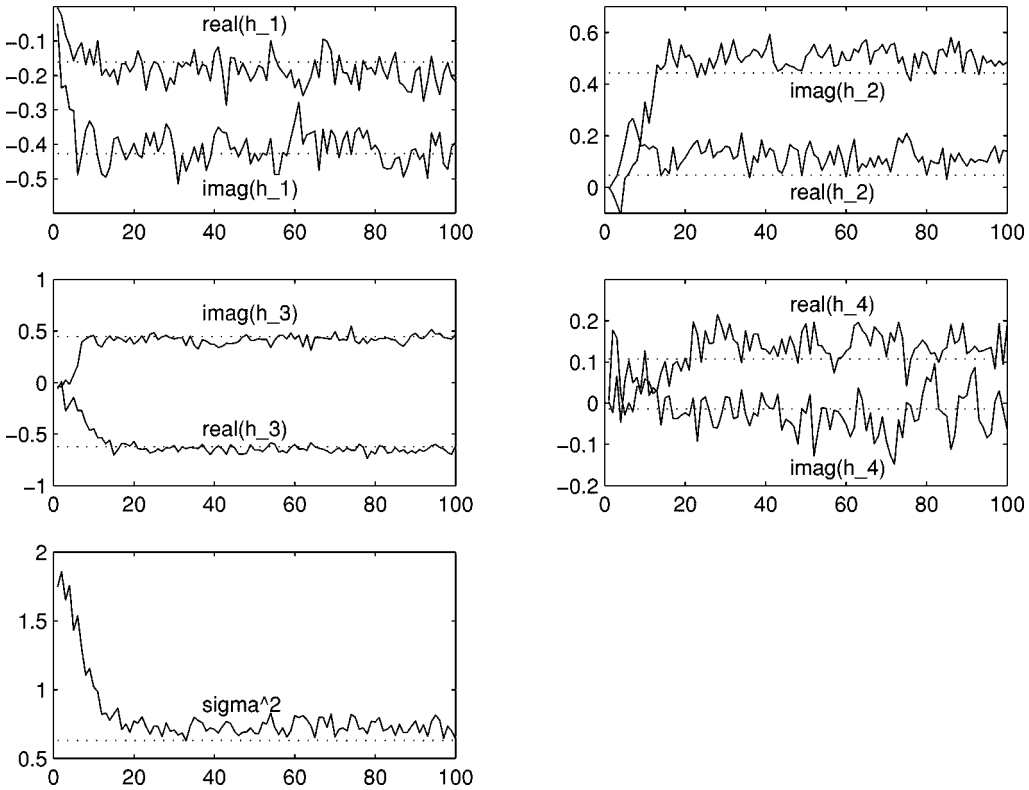

Fig. 3. Samples drawn by the blind Bayesian equalizer in a Gaussian ISI channel. $E_b/N_o = 2$ dB.

The phase ambiguity can be resolved by using differential encoding and decoding [38]. The shift ambiguity can be resolved if we impose constraints on the magnitudes of the channel taps, as will be discussed next. However, note that the use of differential encoding/decoding makes the iterative joint equalization and decoding scheme discussed in Section VI inapplicable. This can be illustrated by the following simple example. Suppose that the code bit sequence is "000 000." In differential encoding, let the first reference bit be "1." Then the transmitted differentially encoded bit sequence becomes "1 111 111." Now suppose that after channel decoding, the decoded code bit sequence is "010010," i.e., decision errors are made on the second and the fifth bits. Then the differentially encoded bit sequence that is fed back to the equalizer (serving as the priors for the code bits) becomes "1 100 011." That is, all the code bits between the two mistaken bits are erroneously encoded. Hence, differential encoding can not be used in systems where the blind turbo equalizer is employed as the receiver.

To resolve the delay and phase ambiguities, we adopt the *constrained* Gibbs sampler along the lines of [11], [12]. For example, we may impose constraints on the phase and amplitude of a particular coefficient, say h_{i_0}, e.g., $\alpha \leq \angle h_{i_0} \leq \beta$, and $|h_{i_0}| \geq \eta$ for some predetermined constants $0 < \alpha < \beta < 2\pi$ and $\eta > 0$. To draw samples of $\boldsymbol{h}$ that satisfy this condition, the so-called *rejection method* [44] can be used. For instance, after a sample is drawn from (15), check to see if the constraint is satisfied; if not, the sample is rejected and a new sample is drawn from (15); the procedure continues until a sample is obtained that satisfies the constraint. If a desired sample has not been obtained after a certain number of rejections, it is more appropriate to *shift* the h_is in the last sample until the i_0th coefficient satisfies the constraint; the vacancies left at the end can be filled with zeros. Another plausible restriction on the amplitudes of h_is is to specify the location of the largest one. For example, we may require that $|h_{i_0}| \geq |h_i| + \eta$ for $i \neq i_0$ and some $\eta > 0$.

2) Initial Synchronization: In order to obtain the above-mentioned constraints on the channel coefficients, the receiver must be synchronized with the transmitter first. This can be accomplished by transmitting a short known *coded* sequence for sounding the channel. At the synchronization stage, upon receiving the transmitted sounding signal, the blind Bayesian equalizer is employed to produce a possibly delayed and phase shifted version of the transmitted symbol sequence. For each possible delay and phase shift, the corresponding code bit sequence is constructed and this sequence is passed through a Viterbi decoder. Each decoded bit sequence is then compared against the original sounding sequence. By locating the best match we can identify the delay and phase ambiguities.

3) Decoder-Assisted Convergence Assessment: Detecting convergence in the Gibbs sampler is usually done in an *ad hoc* way. Some methods can be found in [41]. 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 blind equalizer is followed by a channel decoder, we can assess convergence by monitoring the number of bit corrections made by the channel decoder. If this number exceeds some predetermined threshold, then we decide convergence is not achieved. In that case the Gibbs blind equalizer 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 equalization should be relatively small. On the other hand, if convergence

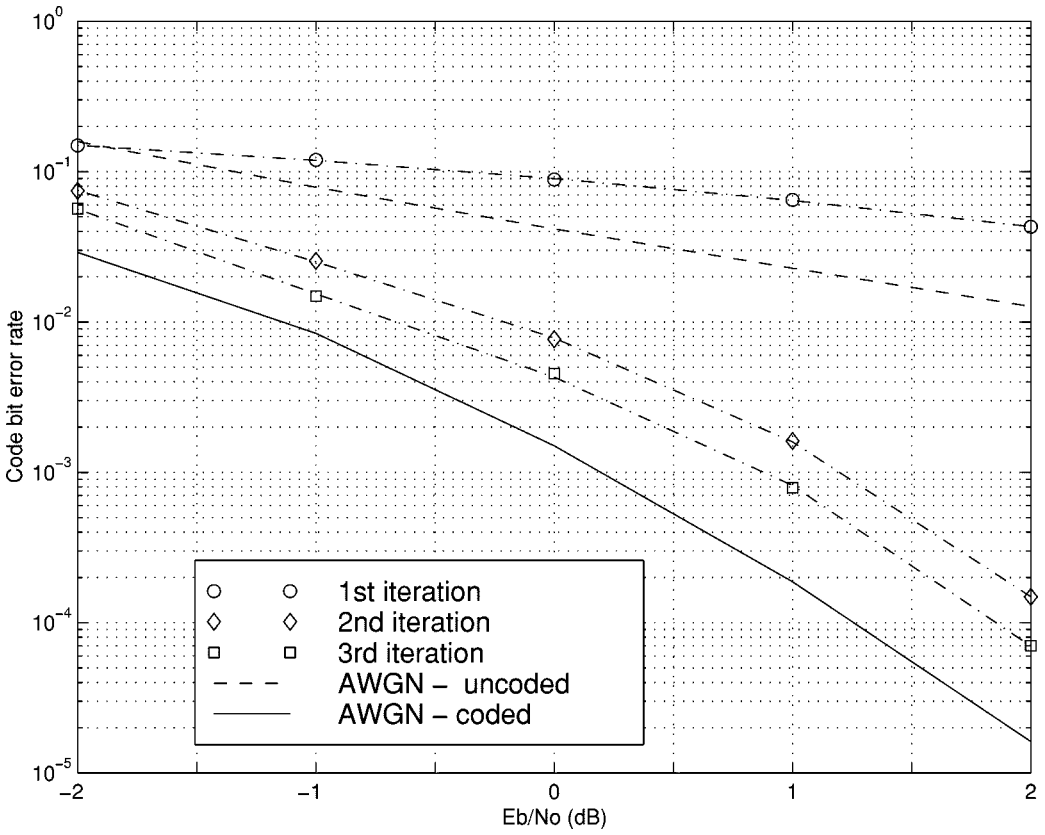

Fig. 4. BER performance of the blind turbo equalizer in a Gaussian ISI channel.

is not reached, then the code bits generated by the equalizer are virtually random and do not satisfy the constraints imposed by the code trellis. Hence, the channel decoder 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 and to count the number of corrections made by the decoder.

4) Relationship Between the Gibbs Sampler and the Expectation-Maximization (EM) Algorithm: The expectation-maximization (EM) algorithm has also been applied to joint channel estimation and equalization in Gaussian [27] and impulsive ISI channels [29]. The major advantage of the Gibbs sampler 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 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 "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 [23] and [46].

VIII. Simulations

In this section, we provide simulation examples to illustrate the performance of the blind Bayesian equalizers developed in this paper. We consider a four-tap ISI channel with complex tap coefficients

$$\boldsymbol{h} = [-0.1611 - j0.4270,\ 0.0467 + j0.4429 \\ -0.6204 + j0.4436,\ 0.1072 - j0.0140]^T.$$

(Note that the channel is normalized to have unit norm, i.e., $|\boldsymbol{h}| = 1$.) In order to resolve the delay and phase ambiguities inherent to the blind equalizer, in the Gibbs sampler, we impose the constraints that $|h_3| > |h_l|$ for $l \in \{1, 2, 4\}$, and $\pi/2 < \angle h_3 \leq \pi$. The channel code is a rate $1/2$ constraint length-five convolutional code (with generators 23, 35 in octal notation). The interleaver is generated randomly and fixed for all simulations. The block size of the information bits is 128 (i.e., $M = 256$). The code bits are BPSK modulated, i.e., $x_k \in \{+1, -1\}$. 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 (22) with $n_0 = N = 50$.

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,

$$p\left(\boldsymbol{h}^{(0)}\right) \sim \mathcal{N}(\boldsymbol{h}_0, \boldsymbol{\Sigma}_0) \longrightarrow \boldsymbol{h}_0 = [0\,0\,0\,0]^T, \quad \boldsymbol{\Sigma}_0 = 1000\,\boldsymbol{I}$$

$$p\left(\sigma^{2(0)}\right) \sim \chi^{-2}(\nu_0, \lambda_0) \longrightarrow \nu_0 = 1, \quad \lambda_0 = 0.1$$

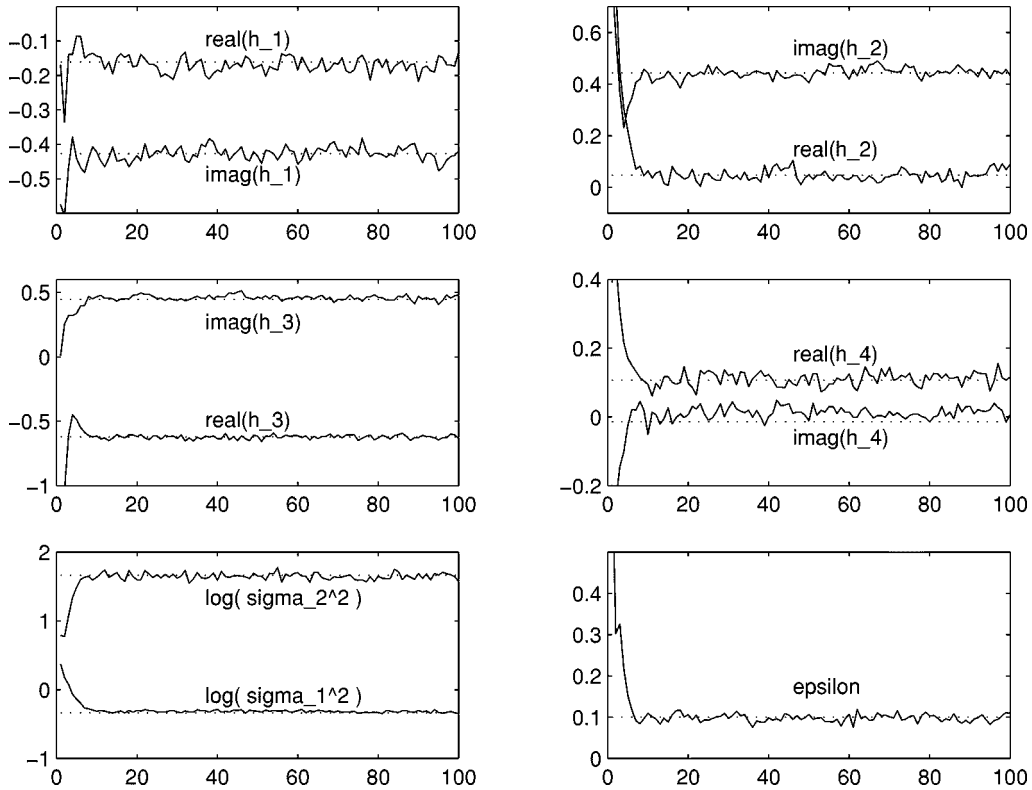

Fig. 5. Samples drawn by the blind Bayesian equalizer in a impulsive ISI channel. $E_b/N_o = -7$ dB.

and for the case of impulsive noise

$$p\left(\boldsymbol{h}^{(0)}\right) \sim \mathcal{N}(\boldsymbol{h}_0, \boldsymbol{\Sigma}_0) \longrightarrow \boldsymbol{h}_0 = [0\,0\,0\,0]^T, \quad \boldsymbol{\Sigma}_0 = 1000\boldsymbol{I}$$

$$p\left(\sigma_1^{2(0)}\right) \sim \chi^{-2}(\nu_1, \lambda_1) \longrightarrow \nu_1 = 1, \quad \lambda_1 = 0.1$$

$$p\left(\sigma_2^{2(0)}\right) \sim \chi^{-2}(\nu_1, \lambda_1) \longrightarrow \nu_1 = 1, \quad \lambda_1 = 1$$

$$p\left(\epsilon^{(0)}\right) \sim \text{Beta}(a_0, b_0) \longrightarrow a_0 = 1, \quad b_0 = 2.$$

In blind turbo equalization, for the first iteration, the prior symbol probabilities $\rho_k \triangleq P[x_k = +1] = 1/2$ for all symbols; in the subsequent iterations, the prior symbol probabilities are provided by the channel decoder, as given by (44). 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.

We first illustrate the performance of the proposed blind Bayesian equalizer in Gaussian ambient noise. In Fig. 3, the convergence behavior of the Gibbs blind equalizer is illustrated for the noise level $\sigma^2 = -2$ dB. The first 100 samples drawn by the Gibbs sampler for the channel taps (h_1, h_2, h_3, h_4) and the noise variance σ^2 are shown. The corresponding true values of these quantities are also shown in the same figure as the dotted lines. It is seen that the Gibbs sampler reaches convergence rapidly (within about 20 iterations). Fig. 4 illustrates the bit error rate (BER) performance of the blind turbo equalizer discussed in Section VI. The code BER at the output of the blind Bayesian equalizer is plotted for the first three iterations. The curve corresponding to the first iteration is the uncoded BER at the output of the blind Bayesian equalizer. The uncoded and coded BER curves in an additive white Gaussian noise (AWGN) ISI-free channel are also shown in the same figure (as, respectively, the dashed and solid lines). It is seen that by incorporating the extrinsic information provided by the channel decoder as the prior symbol probabilities, the proposed blind equalizer achieves the performance that is close to the receiver performance in an ideal AWGN channel in a few iterations.

Next, we illustrate the performance of the blind Bayesian equalizer in an ISI channel with impulsive ambient noise. The noise samples are generated according to the two-term Gaussian model (4) with $\epsilon = 0.1$ and $\sigma_2^2/\sigma_1^2 = 100$. The convergence behavior of the Gibbs blind equalizer for this case is shown in Fig. 5 for total noise level $\sigma^2 \triangleq (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2 = 7$ dB. The first 100 samples of drawn by the Gibbs sampler for the four channel taps (h_1, h_2, h_3, h_4) and the noise parameters $(\sigma_1^2, \sigma_2^2, \epsilon)$ are shown. It is seen that as in the Gaussian noise case, the Gibbs sampler converges within about 20 samples. The BER performance of the blind turbo equalizer in impulsive noise is illustrated in Fig. 6, where the code BERs at the blind Bayesian equalizer for the first three iterations are shown. The uncoded and coded BER curves in an additive white impulsive noise (AWnGN) ISI-free channel are also shown in the same figure (as the dashed and solid lines, respectively). Interestingly, it is seen that with impulsive ambient noise, the performance of the blind Bayesian equalizer in an ISI channel is actually better than the receiver performance in an ISI-free channel. This is not surprising, since an ISI channel introduces memory to the received signal and the channel essentially serves as a trellis code. When a symbol is hit by a large noise impulse, if the channel is ISI-free, i.e., all received signals are independent, then this

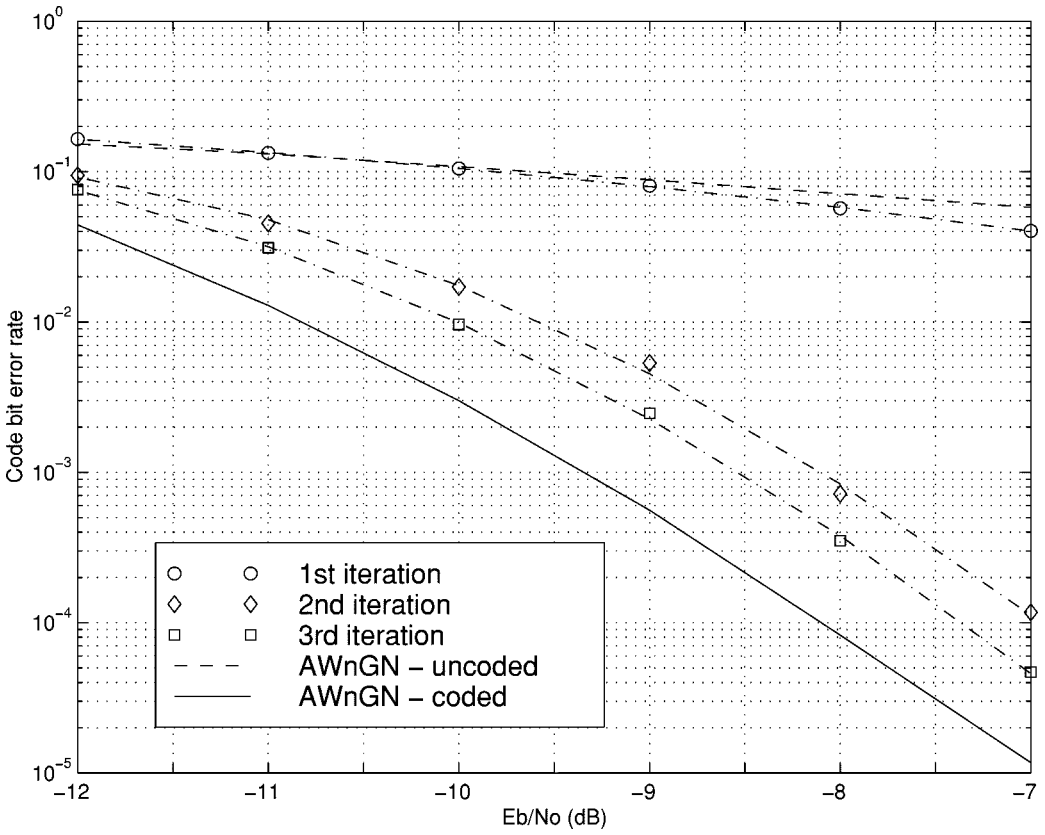

Fig. 6. BER performance of the blind turbo equalizer in a impulsive ISI channel.

symbol cannot be recovered; in an ISI channel, however, by exploiting the channel trellis structure, it is possible to recover this symbol from the adjacent received signals.

IX. Conclusion

In this paper, we have developed a novel blind equalization scheme which is optimal in the sense that it is based on the Bayesian inference of all unknown quantities. Such a blind Bayesian equalizer can be efficiently implemented using the Gibbs sampler, a Markov chain Monte Carlo procedure for computing Bayesian estimates. We have derived the blind Bayesian equalization algorithms for both Gaussian ISI channel and impulsive ISI channel. The proposed blind Bayesian equalizers can incorporate the *a priori* symbol probabilities, and they produces as output the *a posteriori* symbol probabilities. That is, they are "soft-input soft-output" algorithms. Hence, these methods are very well suited for iterative processing in a coded system, which allows the blind Bayesian equalizer to refine its processing based on the information from the decoding stage, and vice versa—a receiver structure termed as *blind turbo equalizer.* Furthermore, the channel decoder facilitates a simple way of assessing the convergence of the blind equalizer by monitoring the number of bit corrections made by the decoder. Finally, we have provided simulations examples to demonstrate the effectiveness of the proposed techniques. It is seen that in both Gaussian and impulsive noise, the blind Bayesian equalizers converge rapidly and the detrimental effect of the ISI channel can be overcome within a few iterations of joint blind equalization and decoding.

Appendix A

Derivation of (15):

$$
\begin{aligned}
&p\left(\boldsymbol{h}|\sigma^2, \boldsymbol{X}, \boldsymbol{Y}\right) \\
&\quad = p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right)}_{\text{not a function of } \boldsymbol{h}} \\
&\quad \propto p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right) \\
&\quad \propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma^2, \boldsymbol{X}\right) p(\boldsymbol{h}) \\
&\quad \propto \exp\left[-\frac{1}{\sigma^2}\sum_{k=0}^{M-1}\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2\right] \\
&\qquad \cdot \exp\left[-(\boldsymbol{h}-\boldsymbol{h}_0)^H\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{h}-\boldsymbol{h}_0)\right] \\
&\quad \propto \exp\left[-\boldsymbol{h}^H\underbrace{\left(\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\boldsymbol{x}_k\boldsymbol{x}_k^H\right)}_{\boldsymbol{\Sigma}_*^{-1}}\boldsymbol{h}\right. \\
&\qquad \left. +2\Re\left\{\boldsymbol{h}^H\underbrace{\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{h}_0 + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\boldsymbol{x}_k y_k^* +\right)}_{\boldsymbol{\Sigma}_*^{-1}\boldsymbol{h}_*}\right\}\right] \\
&\quad \propto \exp\left[-(\boldsymbol{h}-\boldsymbol{h}_*)^H\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{h}-\boldsymbol{h}_*)\right] \sim \mathcal{N}_c(\boldsymbol{h}_*, \boldsymbol{\Sigma}_*). \quad (45)
\end{aligned}
$$

Derivation of (18):

$$
\begin{aligned}
&p\left(\sigma^2|\boldsymbol{h}, \boldsymbol{X}, \boldsymbol{Y}\right)\\
&\quad= p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{h}, \boldsymbol{X}|\boldsymbol{Y}\right)}_{\text{not a function of } \sigma^2}\\
&\quad\propto p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right)\\
&\quad\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma^2, \boldsymbol{X}\right) p(\sigma^2)\\
&\quad\propto \left(\frac{1}{\sigma^2}\right)^{M} \exp\left(-\frac{1}{\sigma^2}\underbrace{\sum_{k=0}^{M-1}\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2}_{s^2}\right)\\
&\qquad\cdot\left(\frac{1}{\sigma^2}\right)^{\nu_0+1} \exp\left(-\frac{\nu_0\lambda_0}{\sigma^2}\right)\\
&\quad= \left(\frac{1}{\sigma^2}\right)^{\nu_0+M+1} \exp\left(-\frac{\nu_0\lambda_0 + s^2}{\sigma^2}\right)\\
&\quad\sim \chi^{-2}\left(2[\nu_0+M],\, \frac{\nu_0\lambda_0+s^2}{\nu_0+M}\right). \qquad (46)
\end{aligned}
$$

Derivation of (20):

$$
\begin{aligned}
&P\left(x_k = a_j|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)\\
&\quad= p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}\big|\boldsymbol{Y}\right)}_{\text{not a function of } x_k}\\
&\quad\propto p\left(\boldsymbol{h}, \sigma^2, \boldsymbol{X}|\boldsymbol{Y}\right)\\
&\quad\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma^2, \boldsymbol{X}\right) p(\sigma^2)\\
&\quad\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma^2, \boldsymbol{X}\right) P(x_k = a_j)\\
&\quad\propto \rho_{kj} \exp\left(-\frac{1}{\sigma^2}\sum_{l=0}^{M-1}\left|y_l - \boldsymbol{h}^H\boldsymbol{x}_l\right|^2\right)\\
&\quad\propto \rho_{kj} \exp\left(-\frac{1}{\sigma^2}\sum_{l=k}^{\overline{k}}\left|y_l - \boldsymbol{h}^H\boldsymbol{x}_l\right|^2\right),\\
&\qquad \text{where } \overline{k} \triangleq \min\{k+Q-1, M-1\} \qquad (47)\\
&\Rightarrow \frac{P\left(x_k = a_j|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}{P\left(x_k = a_i|\boldsymbol{h}, \sigma^2, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}\\
&\quad= \frac{\rho_{kj}}{\rho_{ki}}\\
&\qquad\cdot \exp\left\{-\frac{1}{\sigma^2}\sum_{l=k}^{\overline{k}}\left(\left|y_l - \sum_{m\neq l-k} h_m^* x_{l-m} - h_{l-k}^* a_j\right|^2\right.\right.\\
&\qquad\qquad\left.\left. - \left|y_l - \sum_{m\neq l-k} h_m^* x_{l-m} - h_{l-k}^* a_i\right|^2\right)\right\}\\
&\quad= \frac{\rho_{kj}}{\rho_{ki}} \cdot \exp\left\{-\frac{1}{\sigma^2}\sum_{l=k}^{\overline{k}}\left[|h_{l-k}|^2(|a_j|^2 - |a_i|^2)\right.\right.\\
&\qquad\left.\left. -2\Re\left\{h_{l-k}\left(y_l - \sum_{m\neq l-k} h_m^* x_{l-m}\right)(a_j^* - a_i^*)\right\}\right]\right\}. \qquad (48)
\end{aligned}
$$

APPENDIX B

Derivation of (32):

$$
\begin{aligned}
&p\left(\boldsymbol{h}|\sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)\\
&\quad= p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)}_{\text{not a function of } \boldsymbol{h}}\\
&\quad\propto p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)\\
&\quad\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\right) p(\boldsymbol{h})\\
&\quad\propto \exp\left[-\sum_{k=0}^{M-1}\frac{1}{\sigma_{I_k}^2}\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2\right]\\
&\qquad\cdot \exp\left[-(\boldsymbol{h}-\boldsymbol{h}_0)^H\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{h}-\boldsymbol{h}_0)\right]\\
&\quad\propto \exp\left[-\boldsymbol{h}^H\underbrace{\left(\sum_{k=0}^{M-1}\frac{1}{\sigma_{I_k}^2}\boldsymbol{x}_k\boldsymbol{x}_k^H + \boldsymbol{\Sigma}_0^{-1}\right)}_{\boldsymbol{\Sigma}_*^{-1}}\boldsymbol{h}\right.\\
&\qquad\left. +2\Re\left\{\boldsymbol{h}^H\underbrace{\left(\sum_{k=0}^{M-1}\frac{1}{\sigma_{I_k}^2}\boldsymbol{x}_k y_k^* + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{h}_0\right)}_{\boldsymbol{\Sigma}_*^{-1}\boldsymbol{h}_*}\right\}\right]\\
&\quad\propto \exp\left[-(\boldsymbol{h}-\boldsymbol{h}_*)^H\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{h}-\boldsymbol{h}_*)\right] \sim \mathcal{N}_c(\boldsymbol{h}_*, \boldsymbol{\Sigma}_*). \qquad (49)
\end{aligned}
$$

Derivation of (36):

$$
\begin{aligned}
&p\left(\sigma_i^2|\boldsymbol{h}, \sigma_{\bar{i}}^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)\\
&\quad= p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{h}, \sigma_{\bar{i}}^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)}_{\text{not a function of } \sigma_i^2}\\
&\quad\propto p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)\\
&\quad\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\right) p(\sigma_i^2)\\
&\quad\propto \left(\frac{1}{\sigma_i^2}\right)^{n_i} \exp\left(-\frac{1}{\sigma_i^2}\underbrace{\sum_{k=0}^{M-1}\left|y_k - \boldsymbol{h}^H\boldsymbol{x}_k\right|^2 \mathbf{1}_{\{I_k=i\}}}_{s_i^2}\right)
\end{aligned}
$$

$$\cdot \left(\frac{1}{\sigma_i^2}\right)^{\nu_i+1} \exp\left(-\frac{\nu_i\lambda_i}{\sigma_i^2}\right)$$
$$= \left(\frac{1}{\sigma_i^2}\right)^{\nu_i+n_i+1} \exp\left(-\frac{\nu_i\lambda_i + s_i^2}{\sigma_i^2}\right)$$
$$\sim \chi^{-2}\left(2[\nu_i+n_i],\ \frac{\nu_i\lambda_i+s_i^2}{\nu_i+n_i}\right). \quad (50)$$

Derivation of (37):

$$P\left(x_k = a_j|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)$$
$$= p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}_{[-k]}|\boldsymbol{Y}\right)}_{\text{not a function of } x_k}$$
$$\propto p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)$$
$$\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\right) P(x_k = a_j)$$
$$\propto \rho_{kj} \exp\left(-\sum_{l=0}^{M-1} \frac{1}{\sigma_{I_l}^2} \left|y_l - \boldsymbol{h}^H \boldsymbol{x}_l\right|^2\right)$$
$$\propto \rho_{kj} \exp\left(-\sum_{l=k}^{\overline{k}} \frac{1}{\sigma_{I_l}^2} \left|y_l - \boldsymbol{h}^H \boldsymbol{x}_l\right|^2\right) \quad (51)$$
$$\Rightarrow \frac{P\left(x_k = a_j|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}{P\left(x_k = a_i|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{X}_{[-k]}, \boldsymbol{Y}\right)}$$
$$= \frac{\rho_{kj}}{\rho_{ki}} \cdot \exp\left\{-\sum_{l=k}^{\overline{k}} \frac{1}{\sigma_{I_l}^2} \left(\left|y_l - \sum_{m\neq l-k} h_m^* x_{l-m} - h_{l-k}^* a_j\right|^2 - \left|y_l - \sum_{m\neq l-k} h_m^* x_{l-m} - h_{l-k}^* a_i\right|^2\right)\right\}$$
$$= \frac{\rho_{kj}}{\rho_{ki}} \cdot \exp\left\{-\sum_{l=k}^{\overline{k}} \frac{1}{\sigma_{I_l}^2} \left[|h_{l-k}|^2\left(|a_j|^2 - |a_i|^2\right) - 2\Re\left\{h_{l-k}\left(y_l - \sum_{m\neq l-k} h_m^* x_{l-m}\right)(a_j - a_i)^*\right\}\right]\right\}. \quad (52)$$

Derivation of (38):

$$P\left(I_k = i|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{[-k]}, \boldsymbol{X}, \boldsymbol{Y}\right)$$
$$= p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{[-k]}, \boldsymbol{X}|\boldsymbol{Y}\right)}_{\text{not a function of } I_k}$$
$$\propto p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)$$
$$\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\right) P(I_k = i|\epsilon)$$
$$\propto P(I_k = i|\epsilon) \cdot \frac{1}{\sigma_i^2} \exp\left(-\frac{1}{\sigma_i^2}\left|y_k - \boldsymbol{h}^H \boldsymbol{x}_k\right|^2\right) \quad (53)$$
$$\Rightarrow \frac{P\left(I_k = 1|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{[-k]}, \boldsymbol{X}, \boldsymbol{Y}\right)}{P\left(I_k = 2|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}_{[-k]}, \boldsymbol{X}, \boldsymbol{Y}\right)}$$
$$= \frac{1-\epsilon}{\epsilon} \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \exp\left[\left|y_k - \boldsymbol{h}^H \boldsymbol{x}_k\right|^2 \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right]. \quad (54)$$

Derivation of (39):

$$p\left(\epsilon|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}\right)$$
$$= p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right) \Big/ \underbrace{p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)}_{\text{not a function of } \epsilon}$$
$$\propto p\left(\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}|\boldsymbol{Y}\right)$$
$$\propto p\left(\boldsymbol{Y}|\boldsymbol{h}, \sigma_1^2, \sigma_2^2, \epsilon, \boldsymbol{I}, \boldsymbol{X}\right) p(\epsilon)$$
$$\propto p(\epsilon) p(\boldsymbol{I}|\epsilon)$$
$$\propto \epsilon^{a_0-1}(1-\epsilon)^{b_0-1} \cdot \epsilon^{n_2}(1-\epsilon)^{n_1}$$
$$\sim \text{Beta}(a_0+n_2,\ b_0+n_1). \quad (55)$$

REFERENCES

[1] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, pp. 779–785, May 1970.

[2] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 385–399, June 1980.

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

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correction coding and decoding: Turbo codes," in *Proc. 1993 Int. Conf. Communications (ICC'93)*, Geneva, Switzerland, June 1993, pp. 1064–1070.

[5] 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.

[6] T. K. Blankenship, D. M. Krizman, and T. S. Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," in *Proc. 1997 IEEE Vehicular Technology Conf. (VTC'97)*, Phoenix, AZ, May 1997, pp. 1942–1946.

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

[8] 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.

[9] I. Cha and S. A. Saleem, "Non-linear filtering and equalization in non-Gaussian noise using radial basis function and related networks," in *Proc. 31st Annu. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1997, pp. 13–17.

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

[11] R. Chen and T. H. Li, "Blind restoration of linearly degraded discrete signals by Gibbs sampling," *IEEE Trans. Signal Processing*, vol. 10, pp. 2410–2413, Oct. 1995.

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

[13] K. Dogancay and R. A. Kennedy, "Blind detection of equalization errors in communication systems," *IEEE Trans. Inform. Theory*, vol. 43, pp. 469–482, Mar. 1997.

[14] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glevieux, "Iterative correction of intersymbol interference: Turbo equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sept.–Oct. 1995.

[15] S. Haykin, Ed., *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[16] C. R. Johnson *et al.*, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, pp. 1927–1950, Oct. 1998.
[17] M. Feder and A. Catipovic, "Algorithms for joint channel estimation and data recovery—Application to equalization in underwater communications," *IEEE J. Oceanic Eng.*, vol. 16, pp. 42–55, Jan. 1991.
[18] 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.
[19] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman Hall, 1995.
[20] 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.
[21] M. Ghosh and C. L. Weber, "Maximum likelihood blind equalization," *Opt. Eng.*, vol. 31, pp. 1224–1228, June 1992.
[22] K. Giridhar, J. J. Shynk, and R. A. Iltis, "Bayesian/decision-feedback algorithm for blind equalization," *Opt. Eng.*, vol. 31, pp. 1211–1223, June 1992.
[23] P. J. Green, "Revisible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1985.
[24] J. Hagenauer, "The Turbo principle: Tutorial introduction and state of the art," in *Proc. Int. Symp. Turbo Codes and Related Topics*, Brest, France, Sept. 1997, pp. 1–11.
[25] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for *a priori* unknown channels and symbol timing," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 579–588, Apr. 1992.
[26] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filters," *IEEE Trans. Commun.*, vol. 42, pp. 1017–1032, Mar. 1994.
[27] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, July 1994.
[28] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, Mar. 1985.
[29] R. J. Kozick, R. S. Blum, and B. M. Sadler, "Signal processing in non-Gaussian noise using mixture distribution and the EM algorithm," in *Proc. 31st Annu. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1997, pp. 438–442.
[30] G.-K. Lee, S. B. Gelfand, and M. P. Fitz, "Bayesian techniques for blind deconvolution," *IEEE Trans. Commun.*, vol. 44, pp. 826–835, July 1996.
[31] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2 ed. New York: Springer-Verlag, 1998.
[32] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear process," *Ann. Statistics*, vol. 10, pp. 1195–1208, 1982.
[33] 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.
[34] D. Middleton, "Man-made noise in urban environments and transportation systems: Models and measurement," *IEEE Trans. Commun.*, vol. COM-21, pp. 1232–1241, Nov. 1973.
[35] ——, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. EMC-19, pp. 106–127, 1977.
[36] ——, "Channel modeling and threshold signal processing in underwater acoustics: An analytical overview," *IEEE J. Oceanic Eng.*, vol. OE-12, pp. 4–28, 1987.
[37] D. Middleton and A. D. Spaulding, "Elements of weak signal detection in non-Gaussian noise," in *Advances in Statistical Signal Processing Vol. 2: Signal Detection*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI, 1993.
[38] J. G. Proakis, *Digital Communications*, 3 ed. New York: McGraw-Hill, 1995.
[39] M. J. Schervish and B. P. Carlin, "On the convergence of successive substitution sampling," *J. Computat. Graphical Statist.*, vol. 1, pp. 111–127, 1992.
[40] N. Seshadri, "Joint channel and data estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000–1016, Feb./Mar./Apr. 1994.
[41] M. A. Tanner, *Tools for Statistics Inference*. New York: Springer-Verlag, 1991.
[42] 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.
[43] L. Tong and S. Perreau, "Multichannel blind identification and equalization based on second-order statistics: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, Oct. 1998.
[44] J. von Neumann, "Various techniques used in connection with random digit," *Nat. Bureau Standards Appl. Math. Ser.*, vol. 12, pp. 36–38, 1951.
[45] X. Wang and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, July 1999.
[46] W. H. Wong and F. Liang, "Dynamic importance weighting in Monte Carlo and optimization," *Proc. Nat. Acad. Sci.*, vol. 94, pp. 14220–14 224, 1997.
[47] S. M. Zabin and H. V. Poor, "Efficient estimation of the class A parameters via the EM algorithm," *IEEE Trans. Inform.*, vol. 37, pp. 60–72, Jan 1991.

Xiaodong Wang (M'00) received the B.S. degree in electrical engineering and applied mathematics (highest honor) from Shanghai Jiao Tong University, Shanghai, China, in 1992, the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1995, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1998.

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

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

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.

From September 1990 to May 1999, he was with the Department of Statistics, Texas A&M University, College Station, first as an Assistant Professor and then as an Associate Professor. He is now a Professor of statistics in the Department of Information and Decision Sciences, University of Illinois at Chicago. 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.