BLIND BAYESIAN MULTIUSER DETECTION FOR IMPULSE RADIO UWB SYSTEMS

by

Ahmet Gökhan Ay

B.S., in Electrical and Electronics Engineering, Boğaziçi University, 2003

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical and Electronics Engineering
Boğaziçi University

2006

ACKNOWLEDGEMENTS

I would like to thank Prof. Hakan Deliç and Assist. Prof. Mutlu Koca for their continuous support during this difficult process. I would not be able to complete this thesis without their diligent advisements and encouragements.

My valuable thanks go to my family for their limitless tolerance. I have realized how lucky I am, seeing their indulgence to whatever I do.

Also, I would like to thank to my friends Barış Özgül, Celal Eşli and Nazlı Güney, the people in Wireless Communications Laboratory, for the contribution and joy they have provided during my study.

The research in this thesis was supported by TÜBİTAK EEEAG under Project No. 105E034.

ABSTRACT

BLIND BAYESIAN MULTIUSER DETECTION FOR IMPULSE RADIO UWB SYSTEMS

In this thesis, we address the blind parameter estimation and multiuser detection problems for impulse radio ultra-wide band (UWB) systems under frequency selective fading. The frequency selective UWB channel, having a large number of taps combined with the noise effect, causes certain problems in an UWB system. This channel is often subject to non-Gaussian noise, and with frequency selectivity, there are catastrophic effects. An impulse radio UWB transmitter with orthogonal spreading instead of replication along frames is proposed. In this system, the mentioned blind parameter estimation and multiuser detection problems are considered for three different noise types, Gaussian, impulsive and Cauchy noise. One of the most popular Bayesian estimators, Gibbs sampler, is employed at the receiver part for calculating the Bayesian estimates of the unknown parameters. Because a Gibbs sampler is a soft-input soft-output module capable of exchanging probabilistic information, the proposed detector is also employed within a turbo multiuser detection structure for coded UWB systems. Using this iterative decoding mechanism, further performance improvement is obtained for Gaussian and impulsive noise scenarios, but the sampler fails to converge for the Cauchy case. The bit error rate results for both the Gaussian and impulsive noise cases are given in comparison. Also the tracking performance of the sampler for the unknown parameters are provided for both cases. The simulation results show that the Gibbs sampler is effective in estimating the system parameters and that the proposed receiver provides significant performance gains after a few detection / decoding iterations. It is also concluded that with a proper system design that avoids inter-frame collisions and although an amount of complexity is introduced, the system is suitable for tight bit error rate requirements.

ÖZET

PEK-GENİŞ BANT DÜRTÜ RADYO SİSTEMLER İÇİN GÖZÜ KAPALI ÇOKLU KULLANICI SEZİMİ

Bu tezde, pek-geniş bant (PGB) dürtü radyo (DR) sistemleri için gözü kapalı parametre kestirimi ve çoklu kullanıcı sezimi problemleri ele alınmıştır. Sıklık seçimli PGB kanalı, çok sayıda kanal katsayısı ile tanımlanmıştır. Bu yapısı, PGB sistemlerinde belirli sorunlara yol açar. Bu kanal genelde Gauss olmayan gürültüye maruz kalır ve sıklık seçimlilikle birlikte olumsuz etkiler ortaya çıkarır. Bölümler arasında tekrarlar yerine dikey kodlarla yayma yapılmış bir DR PGB vericisi önerilmiştir. Bu sistemde, bahsi geçen gözü kapalı parametre kestirimi ve çoklu kullanıcı sezimi problemleri üç farklı gürültü modeli için düşünülmüştür : Gauss, dürtün ve Cauchy gürültü. En popüler Bayes kestiricilerinden biri, Gibbs örnekleyicisi, alıcı tarafında bilinmeyen parametreler için Bayes kestirimlerinin hesaplanması amacıyla kullanılmıştır. Gibbs örnekleyicisi, yumuşak-girdi yumuşak-çıktı bir birim olduğundan, önerilen alıcı, kodlu PGB sistemi için turbo çoklu kullanıcı sezim yapısında kullanılmıştır. Bu özyineli kodçözücü mekanizmada, Gauss ve dürtün gürültü senaryolarında performans ilerlemesi elde edilmiş, Cauchy durumunda ise örnekleyici yakınsamamıştır. Hem Gauss, hem de dürtün gürültü senaryoları için hata sonuçları karşılaştırmalı olarak verilmiştir. Ayrıca, Gibbs örnekleyicisinin her iki gürültü tipi için de bilinmeyen parametreleri izleme performansı verilmiştir. Simülasyon sonuçları, Gibbs örnekleyicisinin sistem parametrelerini sezmede etkili olduğunu ve önerilen alıcının birkaç özyineleme sonrasında belirgin performans kazanımı sağladığını göstermektedir. Ayrıca, PGB sisteminde bölümler arasında çarpışmalar olmasını engelleyecek uygun bir tasarımla, bir miktar karmaşıklık eklenmesine rağmen, sistemin sıkı hata gereksinimleri için uygun olduğu dile getirilmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

$\mathbf{A}$	Frequency selective channel matrix
$c_k(i)$	Pseudorandom time hopping sequence
$\mathbf{H}$	Orthogonal spreading matrix
K	Number of all users
M	Number of observed symbols per user
N_c	Number of chips per frame
N_f	Number of frames per symbol
N_h	Number of allowed hopping bins in each frame
n(i)	Noise
$s(t)$	Ultra wide-band signal
T_c	Chip duration
T_f	Frame time
x(i)	Binary information
y(i)	Received binary information
$z(i)$	Received signal from all active users
β	Beta distribution
χ^2	Chi-square distribution
ϵ	Impulse probability
κ	Impulsiveness parameter
μ	Mean
ω	Ultra wide-band waveform
σ^2	Variance
τ	Binary pulse position modulation shift
θ	Vector of unknown parameters
CDMA	Code division multiple access
EM	Expectation maximization
LLR	Log-likelihood ratio

MAP	Maximum a posteriori probability
MCMC	Markov Chain Monte Carlo
MF	Matched filter
MLSD	Maximum-likelihood sequence detector
MUD	Multiuser detection
PPM	Pulse position modulation
SS	Spread spectrum
TH	Time hopping
UWB	Ultra-wide band

1. INTRODUCTION

In an UWB system, a signal is transmitted with a bandwidth much larger than the data modulation bandwidth and thus with a reduced power spectral density. This approach has the potential to produce a signal that is more covert, has higher immunity to interference effects, and has improved time-of-arrival resolution. The spread spectrum (SS) radio system described here is unique in another regard: it does not use a sinusoidal carrier to raise the signal to a frequency band in which signals propagate well, but instead communicates with a time-hopping (TH) baseband signal composed of subnanosecond pulses (referred to as monocycles). An idealized received monocycle shape for a free-space (ignoring multipath, dispersive effects, etc.) channel model is shown in Figure 1.1.

Impulse radio, a form of UWB signaling, has properties that make it a viable candidate for short-range communications in dense multipath environments [1, 2]. Here baseband pulses of very short duration are used for transmission, spreading the energy of the radio signal from near dc to a few gigahertz. Since the bandwidth ranges from near dc to several gigahertz, this impulse radio signal undergoes distortions in the propagation process even in benign propagation environments. On the other hand, the fact that an impulse radio system operates in the lowest possible frequency band that supports its wide transmission bandwidth means that this radio has the best chance of penetrating materials that tend to be more opaque at higher frequencies. Finally, it should be noted that the use of signals with gigahertz bandwidths means that multipath is resolvable down to path differential delays on the order of a nanosecond or less, i.e., down to path length differentials on the order of a foot or less. This significantly reduces fading effects even in indoor environments [3, 4].

The capability to highly resolve multipath combined with the ability to penetrate through materials makes impulse technology viable for high-quality, fully mobile short-range indoor radio systems. Lack of significant multipath fading may considerably reduce fading margins in link budgets and allow low transmission- power operation.

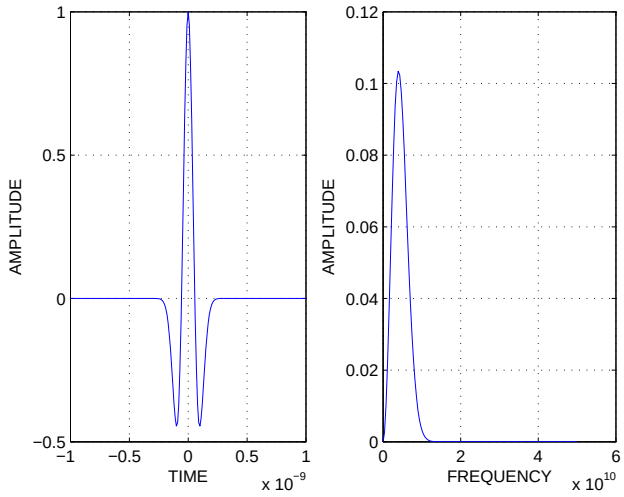

Figure 1.1. 2nd derivative doublet (PPM with 5 pulses) with the expression $y = A * (1 - 4 * pi. * ((t)/\tau).^2). * exp(-2 * pi. * ((t)/\tau).^2), \tau = 0.2$ nanoseconds.

Low transmission-power and short-range operation with UWB results in an extremely low transmitted power spectral density, which insures that impulse radios do not interfere with narrow-band radio systems operating in dedicated bands [2].

Modulation of TH-SS impulse radio is accomplished through the time shifting of pulses. Antipodal modulation cannot be achieved by this means because pulse inversion is not an option in this signaling format [2].

As mentioned, impulse radios make use of SS techniques. A simple way of spreading the spectrum of these UWB low-duty cycle pulse trains is time hopping, with data modulation accomplished by pulse position modulation at the rate of many pulses per data symbol [2]. The key motivations for using TH-SS impulse radio are the ability to highly resolve multipath and the availability of the technology to implement and generate UWB signals with relatively low complexity.

Perhaps UWB communications is more readily known to the radar community under its time-domain description as "baseband carrierless short pulse" techniques. A

comprehensive reference of early work in this area can be found in [5].

The problem of the parameter estimation and data restoration in the presence of channel distortions and additive noise is an active research area. The optimal multiuser detection algorithms with known channel parameters, that is, the multiuser maximum-likelihood sequence detector (MLSD) and the multiuser maximum a posteriori symbol probability (MAP) detector, were first investigated in [6] and [7]. When the channel parameters (e.g., the received amplitudes and the noise variance) are unknown, it is of interest to study the problem of joint multiuser channel parameter estimation and data detection from the received waveform. This problem was first treated in [8], where a solution based on the expectation-maximization (EM) algorithm is derived. In [9], 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 [10], a tree-search algorithm is given for joint data detection and amplitude estimation. Other works concerning multiuser detection with unknown channel parameters include [11]-[16].

Gaussian noise performance of code division multiple access (CDMA) systems are examined for a long time now. The number of studies are well below that of Gaussian for impulsive case and nearly none for Cauchy case, although in many physical channels where multiuser detection may be applied, such as urban and indoor radio channels and underwater acoustic channels, 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 [17]. The results of an early study of error rates in non-Gaussian CDMA channels are found in [18]-[20] in which the performance of the conventional and modified conventional detectors is shown to depend significantly on the shape of the ambient noise distribution. In this thesis, we seek the impulsive noise and Cauchy noise performance of the UWB system as well as a blind estimation technique for unknown parameters.

Of particular interest is the Gibbs sampling approach that is proven to be asymptotically optimum in the estimation performance. A detailed information about

the Gibbs sampler can be found in [21]-[24]. It has been successfully applied to multiuser detection for CDMA [17] and blind equalization of multipath fading channels [25]. Adaptive Bayesian multiuser detectors based on the Gibbs sampler are derived for the Gaussian and impulsive noise synchronous CDMA channels in multiuser detection studies for CDMA. A salient feature of the proposed adaptive Bayesian multiuser detectors is that they can incorporate the *a priori* symbol probabilities, and they produce the a posteriori symbol probabilities [17]. For this reason, we consider its application to a multiuser impulse radio UWB transmission system with parametric uncertainty. Within the parametric uncertainty title, we examine the performance of the system under Gaussian, impulsive and Cauchy noise. But as expected, there are studies regarding the performance of multi-access UWB systems, and very few of them chooses the way of blind parameter estimation [26].

The UWB systems can be observed from a multiuser detection (MUD) point of view. The signal model for a synchronous UWB system is described such that the application of Gibbs parameter estimation is possible. The Gibbs sampler is employed to calculate the Bayesian estimates of the unknown parameters.

In this thesis, the Gibbs sampling procedure is applied to the synchronous UWB model with unknown noise parameters and unknown multipath amplitudes. The log-likelihood ratios of these unknown parameters will be iteratively decoded. A detailed discussion for Turbo codes can be found in [27]-[33]. In [34]-[36], Turbo multiuser detection schemes for coded CDMA systems are proposed, which iterate between multiuser detection and channel decoding to successively improve the receiver performance. The algorithms are adapted to the UWB system. Bayesian inference of all unknown quantities (e.g., the received amplitudes, the data symbols, and the noise parameters) from the received waveforms will be considered, therefore making the algorithm blind.

2. SYSTEM MODEL

We consider the time-hopping multiple access UWB system shown in Figure 2.1, where the kth user signal is expressed as

$$s^{(k)}(t) = \sum_{i=-\infty}^{\infty} \omega\left(t - iT_f^{(k)} - c_k(i)T_c - \tau_m^{(k)}(i)\right) \tag{2.1}$$

where ω is the transmitted monocycle waveform. $T_f^{(k)}$ is the frame time, defined as the time interval that a replica of the symbol is placed in successive frames with a total of N_f frames per symbol. $c_k(i)$ is the time hopping sequence which is selected to be pseudorandom and takes values $0 < c_k(i) < N_h$ where N_h is the number of allowed hopping bins in each frame. T_c is the chip duration, with a total of N_c chips per frame. The monopulse is time-shifted in each frame according to the symbol value; e.g., it is shifted by $\tau_m^{(k)}(i)$ if $x_k\lfloor(\frac{i}{N_f})\rfloor = \mathrm{m}$, $\forall(m) \in \{0, 1\}$, with $\tau_0^{(k)}(i) = 0$ for our binary model and $\tau_1^{(k)}(i) = T_c/2$.

Equation (2.1) can be expressed as

$$s^{(k)}(t) = \sum_{m=0}^{1} s_m^{(k)}(t) \tag{2.2}$$

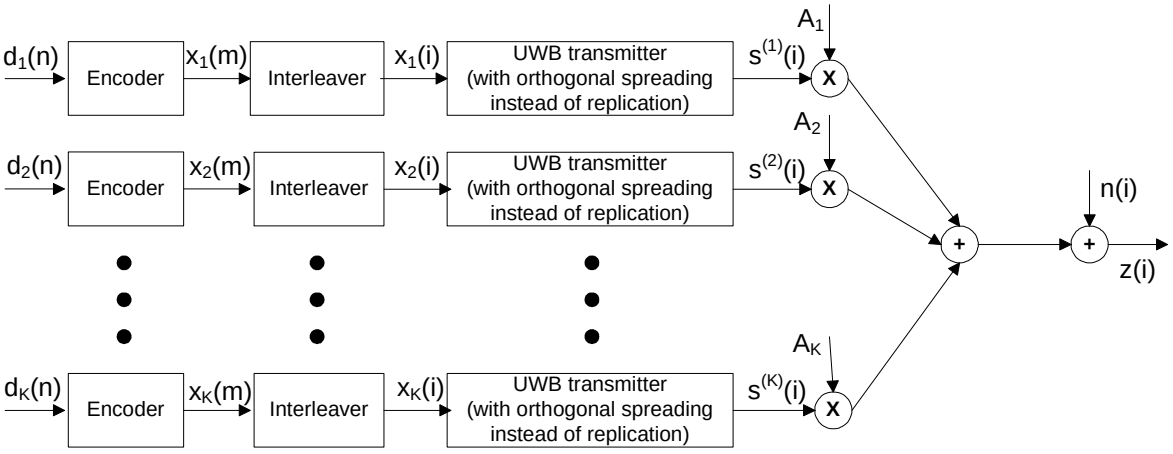

Figure 2.1. The block diagram of the UWB system.

with

$$s_m^{(k)}(t) = \sum_{i=-\infty}^{\infty} \omega\left(t - iT_f^{(k)} - c_k(i)T_c - \tau_m^{(k)}(i)\right) g_m^{(k)}(i) \tag{2.3}$$

where $g_m^{(k)}(i) = \delta(x_k(\lfloor(\frac{i}{N_f})\rfloor) - m)$, $\forall m \in \{0, 1\}$. If we define $\omega_m(t) = \omega(t - \tau_m^{(k)}(i))$, we can rewrite (2.3) as

$$s_m^{(k)}(t) = \sum_{i=-\infty}^{\infty} \omega_m\left(t - (iN_c + c_k(i))T_c\right) g_m^{(k)}(i). \tag{2.4}$$

So, $\omega_m(t)$ seems to be linearly modulated with symbol rate $1/T_c$.

The inherent "replication of symbols along the frames" property of the UWB model is recognized to be a kind of spreading. The change is that the conventional UWB system repeats the same symbol along frames while a length-K orthogonal spreading code is used for this purpose in the proposed model, with **H** being the spreading matrix of dimensions $N_f \,\mathrm{x}\, K$, where K is the total number of users as given in Figure 2.1. $\omega(t)$ in (2.1), as mentioned, can be observed to be linearly modulated with symbol rate $1/T_c$ [37]. We can obtain a discrete time model representation by sampling every T_c seconds. Then the information from all users at the receiver will be shown as

$$\mathbf{z}(i) = \mathbf{A}\,\mathbf{s}(i) + \mathbf{n}(i) \tag{2.5}$$

for **s**(i) and **A** defined in the definitions below with **A** being the matrix formed using the coefficients of the frequency selective UWB channel of length K, where the channel corresponding to the kth user is $A_k(t) = \sum_{j=0}^{K-1} \alpha_{kj}\delta(t - \tau_{kj})$, with α's being the gain and δ's the delay respectively [38]. The different paths are assumed to arrive at integer multiples of $\tau_1(i)$. The noise variance(s) and the channel coefficients are assumed to be constant during the observed M symbols per user.

The UWB receiver knows $c_k(i)$'s for all of K users' observed M symbols each. The receiver of the proposed system determines the symbol values at positions in each

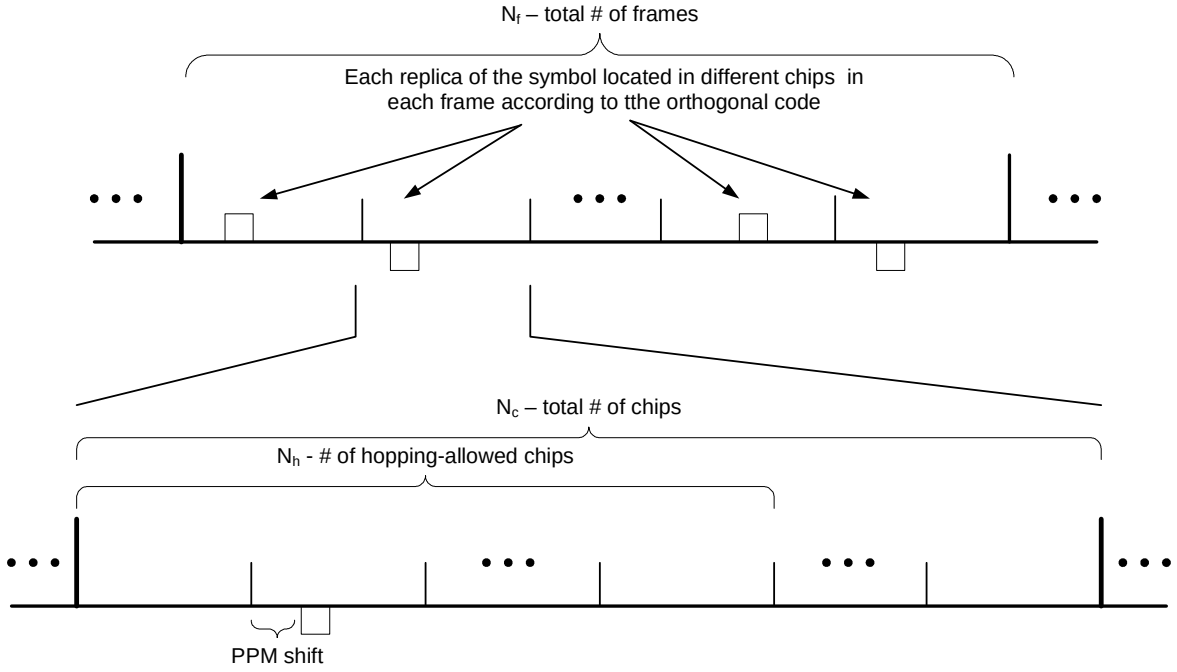

Figure 2.2. The signal model.

frame pointed out by $c_k(i)$'s. Detecting the orthogonally spread and distorted symbol values in each frame of the received waveforms, $\mathbf{z}(i)'s$, allows us to form a matrix, $\mathbf{G}$(i) defined in (2.6), same as the spread bits of a CDMA user in a CDMA system. From this point on, the procedure is symbol value estimation under Gaussian noise and multipath fading in an UWB system with orthogonal spreading applied over successive frames. For computing the values of $\mathrm{P}(\mathrm{x}_k(i) = 1|\mathbf{Y})$ for *k = 1, 2, ..., K* and *i = 0, 1, ..., M-1*, we must first make some definitions:

$$
\begin{aligned}
\mathbf{x}(i) &= [x_1(i)\ x_2(i)\ ...\ x_K(i)]^T \\
\mathbf{s}(i) &= [s^{(1)}(i)\ s^{(2)}(i)\ ...\ s^{(K)}]^T \\
\mathbf{B}(i) &= \mathrm{diag}[x_1(i),\ x_2(i),\ ...,\ x_K(i)] \qquad i = 0, 1, 2, ..., M-1 \\
\mathbf{H} &= [\mathbf{h}_1\ \mathbf{h}_2\ ...\ \mathbf{h}_K] \\
\mathbf{X} &= \{\mathbf{x}(0), \mathbf{x}(1), ..., \mathbf{x}(M-1)\} \\
\mathbf{Y} &= \{\mathbf{y}(0), \mathbf{y}(1), ..., \mathbf{y}(M-1)\} \\
\mathbf{a} &= [A_1\ A_2\ ...\ A_K]^T \\
\mathbf{A} &= \mathrm{diag}[A_1, A_2, ..., A_K].
\end{aligned}
$$

We can easily form the matrix $\mathbf{G}$(i) as

$$\mathbf{G}(i) = \mathbf{HB}(i) \qquad i = 0, 1, ..., M-1 \tag{2.6}$$

from the values detected at chips the $c_k(i)$ values point as explained previously. Then we can rewrite our problem as

$$\begin{aligned} \mathbf{y}(i) &= \mathbf{H}\ \mathbf{A}\ \mathbf{x}(i) + \mathbf{n}(i) && (2.7) \\ &= \mathbf{H}\ \mathbf{B}(i)\ \mathbf{a} + \mathbf{n}(i) \qquad i = 0, 1, ..., M-1. && (2.8) \end{aligned}$$

where $\mathbf{B}(i)$ can easily be formed as

$$\mathbf{B}(i) = \mathbf{H}^{-1}\mathbf{G}(i) \qquad i = 0, 1, ..., M-1. \tag{2.9}$$

Now the problem, estimating user symbols under Gaussian, impulsive or Cauchy noise and UWB channel fading, can be handled using the Gibbs sampler. But first, in the next section, we describe the logic behind the application of Gibbs sampler for symbol estimation under unknown UWB fading channel with Gaussian and impulsive noise. We will briefly discuss Bayesian analysis, so that we can justify our prior selections in the following sections.

3. BAYESIAN ANALYSIS

Gibbs sampling is one of several Markov Chain Monte Carlo (MCMC) methods. The aim of these methods is to take samples from the posterior distribution. They are called Monte Carlo because they involve drawing random numbers from specified distributions and Markov chain because each sample depends on the previous sample. MCMC methods can be used by frequentists but, because they involve sampling from the posterior distribution, they fit more conveniently into the statistical tool box of Bayesians.

The value of MCMC methods is that they can be used in analysis that are very difficult to perform analytically. They work by breaking a very complex problem down into a series of simple steps. Their popularity has grown in recent years because they often require a lot of computer power to generate many samples and this is only practical with fast computers. Gibbs sampling is used widely because it is the most popular MCMC method in genetic analysis but other methods such as the Metropolis-Hastings algorithm are also used [39].

For a better understanding in Gibbs sampling, one shall explore the Bayesian Statistics. As opposed to the point estimators (means, variances) used by classical statistics, Bayesian statistics is concerned with generating the posterior distribution of the unknown parameters given both the data and some prior density for these parameters [40, 41]. As such, Bayesian statistics provides a much more complete picture of the uncertainty in the estimation of the unknown parameters, especially after the confounding effects of nuisance parameters are removed[39]. Our treatment here is intentionally quite brief and just about to provide an insight to the basic ideas employed in this thesis.

3.1. Bayes' Theorem

The foundation of Bayesian statistics is Bayes theorem. Suppose we observe a random variable y and wish to make inferences about another random variable θ, where θ is drawn from some distribution $p(\theta)$. Using Bayes' theorem, $p(\theta|y)$ can be expressed in terms of $p(y|\theta)$ as

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}. \tag{3.1}$$

Similarly with n possible outcomes $(\theta_1, \theta_2, ..., \theta_n)$,

$$p(\theta_j|y) = \frac{p(y|\theta_j)p(\theta_j)}{p(y)} = \frac{p(y|\theta_j)}{\sum_{i=1}^{n} p(b_i)p(y|\theta_i)} \tag{3.2}$$

where $p(\theta)$ is the prior distribution of the possible θ values, while $p(\theta|y)$ is the posterior distribution of θ given the observed data y [39].

3.2. From Likelihood to Bayesian Analysis

Suppose $\ell(\theta|y)$ is the assumed likelihood function and we start with some initial knowledge/guess about the distribution of the unknown parameter(s), $p(\theta)$. From Bayes' theorem, the data (likelihood) augment the prior distribution to produce a posterior distribution,

$$\begin{aligned} p(\theta|y) &= \frac{1}{p(y)}.p(y|\theta).p(\theta) \\ &= (\text{normalizing constant}).p(y|\theta).p(\theta) \\ &= \text{constant x likelihood x prior} \end{aligned} \tag{3.3}$$

as $p(y|\theta) = \ell(\theta|y)$ is just the likelihood function. $1/p(y)$ is a constant (with respect to θ), because our concern is the distribution over θ. Because of this, the posterior

distribution is often written as

$$p(\theta|y) \propto \ell(\theta|y)p(\theta). \tag{3.4}$$

Note that the constant p(y) normalizes $p(y|\theta).p(\theta)$ to one, and hence can be obtained by integration,

$$p(y) = \int_{\theta} p(y|\theta).p(\theta)d\theta. \tag{3.5}$$

The dependence of the posterior on the prior (which can easily be assessed by trying different priors) provides an indication of how much information on the unknown parameter values is contained in the data. If the posterior is highly dependent on the prior, then the data likely has little signal, while if the posterior is largely unaffected under different priors, the data are likely highly informative.

3.2.1. Marginal Posterior Distributions

Often, only a subset of the unknown parameters is really of concern to us, the rest being nuisance parameters that are really of no concern to us. A very strong feature of Bayesian analysis is that we can remove the effects of the nuisance parameters by simply integrating them out of the posterior distribution to generate a marginal posterior distribution for the parameters of interest. For example, suppose the mean and variance of data coming from a normal distribution are unknown, but our real interest is in the variance. Estimating the mean introduces additional uncertainly into our variance estimate. This is not fully capture in standard classical approaches, but under a Bayesian analysis, the posterior marginal distribution for σ^2 is simply

$$p(\sigma^2|y) = \int p(\mu, \sigma^2|y)d\mu. \tag{3.6}$$

The marginal posterior may involve several parameters (generating joint marginal posteriors).Write the vector of unknown parameters as $\theta = [\theta_1 \quad \boldsymbol{\theta}_n]^T$, where $\theta_{\mathbf{n}}$ is the

vector of nuisance parameters. Integrating over $\theta_{\mathbf{n}}$ gives the desired marginal as

$$p(\theta_1|y) = \int_{\theta_{\mathbf{n}}} p(\theta_1, \theta_{\mathbf{n}}|y) d\theta_{\mathbf{n}}. \tag{3.7}$$

3.3. The Choice of a Prior

Obviously, a critical feature of any Bayesian analysis is the choice of a prior. The key here is that when the data have sufficient signal, even a bad prior will still not greatly influence the posterior. In a sense, this is an asymptotic property of Bayesian analysis in that all but pathological priors will be overcome by sufficient amounts of data. As mentioned above, one can check the impact of the prior by seeing how stable to posterior distribution is to different choices of priors. If the posterior is highly dependent on the prior, then the data (the likelihood function) may not contain sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contain significant information.

The location of a parameter (mean or mode) and its precision (the reciprocal of the variance) of the prior is usually more critical than its actual shape in terms of conveying prior information. The shape (family) of the prior distribution is often chosen to facilitate calculation of the prior, especially through the use of conjugate priors that, for a given likelihood function, return a posterior in the same distribution family as the prior (i.e., a gamma prior returning a gamma posterior when the likelihood is Poisson)[39].

3.4. Posterior Distributions Under Normality Assumptions

To introduce the basic ideas of Bayesian analysis, consider the case where data are drawn from a normal distribution, so that the likelihood function for the ith ob-

servation, y_i is

$$\ell(\mu, \sigma^2|y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right). \tag{3.8}$$

The resulting full likelihood for all n data points is

$$\ell(\mu|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left[(\sum_{i=1}^{n} y_i^2) - 2n\bar{y}\mu + n\mu^2\right]\right). \tag{3.9}$$

3.4.1. Known Variance and Unknown Mean

Assume the variance σ^2 is known, while the mean μ is unknown. For a Bayesian analysis, it remains to specify the prior for μ, $p(\mu)$. Suppose we assume a Gaussian prior, $\mu \sim N(\mu_0, \sigma_0^2)$, so that

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right). \tag{3.10}$$

The mean and variance of the prior, μ_0 and σ_0^2 are referred to as hyperparameters. One important trick we will use throughout when calculating the posterior distribution is to ignore terms that are constants with respect to the unknown parameters. Suppose y denotes the data and $\theta_\mathbf{1}$ is a vector of known model parameters, while $\theta_\mathbf{2}$ is a vector of unknown parameters. If we can write the posterior as

$$p(\theta_2|y, \theta_1) = f(y, \theta_1).g(y, \theta_1, \theta_2) \tag{3.11}$$

then,

$$p(\theta_2|y, \theta_1) \propto g(y, \theta_1, \theta_2). \tag{3.12}$$

With the prior given by above equation, we can express the resulting posterior distribution as

$$p(\mu|y) \propto \ell(\mu|y).p(\mu) \propto \exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2} - \frac{1}{2\sigma^2}[\sum_{i=1}^{n} y_i^2 - 2n\bar{y}\mu + n\mu^2]\right). \tag{3.13}$$

We can factor out additional terms not involving μ to give

$$p(\mu|y) \propto \exp\left(-\frac{\mu^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0^2} + \frac{n\bar{y}\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right). \tag{3.14}$$

Factoring in terms of μ, the term in the exponential becomes

$$-\frac{\mu^2}{2}(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}) + \mu(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}) = -\frac{\mu^2}{\sigma_*^2} + \frac{2\mu\mu_*}{2\sigma_*^2} \tag{3.15}$$

where

$$\sigma_*^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1} \quad \text{and} \quad \mu_* = \sigma_*^2\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right). \tag{3.16}$$

Finally, by completing the square, we have

$$p(\mu|y) \propto \exp\left(-\frac{(\mu-\mu_*)^2}{2\sigma_*^2} + f(y, \mu_0, \sigma^2, \sigma_0^2)\right). \tag{3.17}$$

The posterior density function for μ thus becomes

$$p(\mu|y) \propto \exp\left(-\frac{(\mu-\mu_*)^2}{2\sigma_*^2}\right). \tag{3.18}$$

Recalling that the density function for $z \sim N(\alpha, \beta)$ is

$$p(z) \propto \exp\left(-\frac{(z-\alpha)^2}{2\beta}\right) \tag{3.19}$$

shows that the posterior density function for μ is a normal with mean μ_* and variance σ_*^2, e.g.,

$$\mu|(y, \sigma^2) \sim N(\mu_*, \sigma_*^2). \tag{3.20}$$

Notice that the posterior density is in the same form as the prior. This occurred because the prior conjugated with the likelihood function - the product of the prior and likelihood returned a distribution in the same family as the prior. The use of such conjugate priors (for a given likelihood) is a key concept in Bayesian analysis. It is needed to inquire about the relative importance of the prior versus the data. Under the assumed prior, the mean of the posterior distribution is given by

$$\mu_* = \mu_0 \frac{\sigma_*^2}{\sigma_0^2} + \bar{y} \frac{\sigma_*^2}{\sigma^2/n} \tag{3.21}$$

Note with a very diffuse prior on μ (i.e. $\sigma_0^2 >> \sigma^2$), that $\sigma_*^2 \to \sigma^2/n$ and $\mu_* \to y$. Also note that as we collect enough data, $\sigma_*^2 \to \sigma^2/n$ and again $\mu_* \to x$ [39].

3.4.2. Gamma, Inverse-Gamma, χ^2 and χ^{-2} Distributions

Before we examine a Gaussian likelihood with unknown variance, a brief aside is needed to develop χ^{-2}, the inverse chi-square distribution. We do this via the gamma and inverse-gamma distributions.

The χ^2 is a special case of the Gamma distribution, a two parameter distribution. A gamma-distributed variable is denoted by $y \sim$ Gamma (α, β), with density function

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y\beta) \qquad \text{for } \alpha, \beta, y > 0. \tag{3.22}$$

As a function of y, note that

$$p(y) = y^{\alpha-1} \exp(-y\beta). \tag{3.23}$$

We can parameterize a gamma in terms of its mean and variance by noting that

$$\mu_y = \frac{\alpha}{\beta}, \quad \sigma_y^2 = \frac{\alpha}{\beta^2}. \tag{3.24}$$

$\Gamma(\alpha)$, the gamma function evaluated at α is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt \tag{3.25}$$

with t being a dummy variable. The gamma function is the generalization of the factorial function (n!) to all positive numbers.

The χ^2 distribution is a special case of the gamma, with a χ^2 with n degrees of freedom being a gamma random variable with $\alpha = n/2$, $\beta = n/2$, i.e., $\chi_n^2 \sim$ Gamma$(n/2, 1/2)$, giving the density function as

$$p(y|n) = \frac{2^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} y^{(\frac{n}{2})-1} \exp\left(-\frac{y}{2}\right). \tag{3.26}$$

Hence for a χ_n^2

$$p(y) \propto y^{(\frac{n}{2})-1} \exp\left(-\frac{y}{2}\right). \tag{3.27}$$

The inverse gamma distribution will prove useful as a conjugate prior for Gaussian likelihoods with unknown variance. It is defined by the distribution of $t = 1/y$, where $y \sim$ Gamma(α, β). The resulting density function, mean, and variance become

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha-1)} \exp\left(-\frac{\beta}{y}\right) \quad \text{for} \quad \alpha, \beta, y > 0 \tag{3.28}$$

$$\mu_y = \frac{\beta}{\alpha - 1}, \quad \sigma_y^2 = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}. \tag{3.29}$$

Note for the inverse gamma that

$$p(y) \propto y^{-(\alpha-1)} \exp\left(-\frac{\beta}{y}\right). \tag{3.30}$$

If $y \sim \chi_n^2$, then $t = 1/y$ follows an χ_n^{-2} distribution, and denote this by $t \sim \chi_n^{-2}$. This is a special case of the inverse gamma, with (as for a normal χ^2) $\alpha = n/2, \beta = 1/2$. The resulting density function is

$$p(y|n) = \frac{2^{\left(-\frac{n}{2}\right)}}{\Gamma\left(\frac{n}{2}\right)} y^{(1-\frac{n}{2})} \exp\left(-\frac{1}{2y}\right) \tag{3.31}$$

with mean and variance

$$\mu_y = \frac{1}{n-2}, \quad \sigma_y^2 = \frac{2}{(n-2)^2(n-4)}. \tag{3.32}$$

The scaled inverse chi-square distribution is more typically used, where

$$p(y|n) \propto y^{1-(\frac{n}{2})} \exp\left(-\frac{\sigma_0^2}{2y}\right) \tag{3.33}$$

so that the $1/2y$ term in the exponential is replaced by an $\sigma_0^2/2y$ term. If y follows this distribution, then $\sigma_0^2.y$ follows a standard χ^{-2} distribution. The scaled-inverse chi-square distribution thus involves two parameters, σ_0^2 and n and it is denoted by $\chi^{-2}_{(n,\sigma_0^2)}$ [39]. Note that if

$$y \sim \chi^{-2}_{(n,\sigma_0^2)} \quad \text{then} \quad \sigma_0^2 y \sim \chi_n^{-2} \tag{3.34}$$

Table 3.1. Summary of the functional forms of the distributions introduced

Distribution	p(y)/constant
Gamma(α, β)	$y^{\alpha-1}\exp(-\beta y)$
χ^2_n	$y^{(n/2)-1}\exp\left(-\frac{y}{2}\right)$
Inverse-Gamma(α, β)	$y^{-(\alpha-1)}\exp\left(-\frac{\beta}{y}\right)$
χ^{-2}_n	$y^{1-n/2}\exp\left(-\frac{1}{2y}\right)$
$\chi^{-2}_{n,S}$	$y^{1-n/2}\exp\left(-\frac{S}{2y}\right)$

3.4.3. Unknown Variance: χ^{-2} Priors

Now suppose the data are drawn from a normal with known mean μ, but unknown variance σ^2. The resulting likelihood function becomes

$$\ell(\sigma^2|y,\mu) \propto (\sigma^2)^{-n/2}.\exp\left(-\frac{nS^2}{2\sigma^2}\right) \tag{3.35}$$

where

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2. \tag{3.36}$$

Notice that since we condition on y and μ (i.e., their values are known), the S^2 is a constant. Further observe that, as a function of the unknown variance σ^2, the likelihood is proportional to a scaled χ^{-2} distribution. Thus, taking the prior for the unknown variance also as a scaled χ^{-2} with hyperparameters ν_0 and σ_0^2, the posterior becomes

$$\begin{aligned} p(\sigma^2|y,\mu) &\propto (\sigma^2)^{-n/2}\exp\left(\frac{-nS^2}{2\sigma^2}\right)(\sigma^2)^{-1-\nu_0/2}\exp\left(-\frac{\sigma_0^2}{2\sigma^2}\right) \\ &= (\sigma^2)^{-1-(n+\nu_0)/2}\exp\left(-\frac{nS^2+\sigma_0^2}{2\sigma^2}\right). \end{aligned} \tag{3.37}$$

This is also a scaled χ^{-2} distribution with parameters $\nu_0 = (n+\nu_0)$ and $\sigma_n^2 = (nS^2+\sigma_0^2)$, so that

$$\sigma_n^2\sigma^2|(y,\mu) \sim \chi_{\nu_n}^{-2}. \tag{3.38}$$

3.5. Conjugate Priors

The use of a prior density that conjugates the likelihood allows for analytic expressions of the posterior density. Table 2 gives the conjugate priors for several common likelihood functions.

Table 3.2. Conjugate priors for common likelihood functions

Likelihood	Type	Conjugate prior
Binomial	-	Beta
Multinomial	-	Dirichlet
Poisson	-	Gamma
Normal	μ unknown, σ^2 known	Normal
Normal	μ known, σ^2 unknown	Inverse-Chi Square
Multivariate Normal	μ unknown, V known	Multivariate Normal
Multivariate Normal	μ known, V unknown	Inverse Wishart

3.6. Gibbs Sampler Overview

The Gibbs sampler works on Bayesian estimation principles for solving the estimation problem of the entire posterior density of the parameters having high dimensions [42]. The Gibbs sampler works as described below.

Let $\theta = [\theta_1 \dots \theta_d]^T$ be a vector of unknown parameters, and let $\mathbf{Y}$ 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 $\mathbf{Y}$, i.e.,$p(\theta_j|\mathbf{Y})$, $1 \leq j \leq d$.

Direct evaluation involves integrating out the rest of the parameters from the joint a posteriori density, i.e.,

$$p(\theta_j|\mathbf{Y}) = \int\int \ldots \int p(\theta|\mathbf{Y}) d\theta_1 \ldots d\theta_{j-1} d\theta_{j+1} \ldots d\theta_d. \tag{3.39}$$

In most cases, such a direct evaluation is computationally infeasible, especially when the parameter dimension d is large. The Gibbs sampler is a Monte Carlo procedure for numerical evaluation of the above multidimensional integral. The basic idea is to generate random samples from the joint posterior distribution $p(\theta|\mathbf{Y})$ and then to estimate any marginal distribution using these samples. Given the initial values $\theta^{(0)} = [\theta_1^{(0)} \ldots \theta_d^{(0)}]^T$, 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* $p(\theta_d|\theta_1^{(n+1)}, \ldots, \theta_{d-1}^{(n+1)}, \boldsymbol{Y})$.

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

$$K(\theta^{'}, \theta) = p(\theta_1|\theta_2^{'}, \ldots, \theta_d^{'}, \mathbf{Y}) p(\theta_2|\theta_1, \theta_3^{'}, \ldots, \theta_d^{'}, \mathbf{Y}) \ \ldots \ p(\theta_d|\theta_1, \ldots, \theta_{d-1}, \mathbf{Y}). \tag{3.40}$$

The convergence behavior of the Gibbs sampler has the general conditions that are given for the following two results:

1. The distribution of $\theta^{(n)}$ converges geometrically to $p(\theta|\mathbf{Y})$ as $n \to \infty$.
2. $\frac{1}{N} \sum_{n=1}^{N} f(\theta^{(n)}) \to \int f(\theta) p(\theta|\mathbf{Y}) d\theta$, as $N \to \infty$, for any integrable function f.

The Gibbs sampler requires an initial transient period to converge to equilibrium. The initial period of length n_0 is known as the "burn-in" period, and the first n_0 samples

should always be discarded. Detecting convergence is usually done in some ad hoc way. This results in hard experiences in deciding when to make the last iteration before convergence. This can be solved by selecting a long enough period for the burn-in.

4. BLIND BAYESIAN MULTIUSER DETECTION FOR UWB SYSTEMS

In this section we will present the operation of Gibbs sampler for Gaussian, impulsive and lastly for Cauchy noise models.

4.1. Multiuser Detection in Gaussian Noise

In our analysis, the Gibbs sampler is employed as described below.

We have a vector $\theta = [\mathbf{a} \;\; \sigma^2 \;\; \boldsymbol{X}]^T$ of unknown parameters each of which is defined in 4.1.1., and $\mathbf{Y}$ is the observed data. We are interested in finding the a posteriori marginal distribution of some parameter, say σ^2, conditioned on the observation $\mathbf{Y}$, i.e., $p(\sigma^2|\mathbf{Y})$. In this case, direct evaluation involves

$$p(\sigma^2|\mathbf{Y}) = \int \int \ldots \int p(\theta|\mathbf{Y})\, d\mathbf{a}\, d\boldsymbol{X} \tag{4.1}$$

for all the elements of $\mathbf{a}$ and $\boldsymbol{X}$ and with

$$\begin{aligned} p(\theta|\mathbf{Y}) &= p(\mathbf{Y}|\mathbf{a}, \sigma^2, \mathbf{X}|)\; p(\mathbf{a})\; p(\sigma^2)\; p(\mathbf{X})/p(\mathbf{Y}) && (4.2) \\ &= C\left(\frac{1}{\sigma^2}\right)^{PM/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{M-1} \|\mathbf{y}(i) - \mathbf{HAx}(i)\|^2\right) p(\mathbf{a})\; p(\sigma^2)\; p(\mathbf{X}) && (4.3) \end{aligned}$$

where C is a normalization constant independent of the unknown parameters. Because the computation for integrating out the rest of the parameters to obtain the a posteriori probability of the desired parameter involves 2^{KM-1} multidimensional integrals with K and M defined in chapter 2, being the total number of active users and number of symbols per user respectively, the Gibbs sampler is used to generate random samples from the joint posterior distribution $p(\theta|\mathbf{Y})$ and then to estimate any marginal distribution using these samples [42].

4.1.1. Prior Distributions

Bayesian analysis requires a careful selection of prior distributions. The following are the basic assumptions due to the considerations of [17]:

1. The ambient noise distribution is Gaussian. The pdf of n(i) is given by

$$p(n(i))\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{\|n(i)\|^2}{2\sigma^2}\right). \tag{4.4}$$

2. For the unknown amplitude vector $\mathbf{a} = [A_1\, A_2\, ...\, A_K]^T$ where A_k's are the channel coefficients, a truncated Gaussian prior distribution is assumed with

$$p(\mathbf{a}) \propto N\left(\mathbf{a}_0, \mathbf{\Sigma}_0\right) I_{a>0} \tag{4.5}$$

   where $I_{a>0}$ is an indicator that is 1 if all the elements of $\mathbf{a}$ are positive and is 0 otherwise. This means, the algorithm can be used for positive channel amplitudes only.

3. For the noise variance σ^2, a χ^{-2} prior distribution is assumed

$$p(\sigma^2) \;=\; \frac{\left(\frac{\nu_0\lambda_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)}\left(\frac{1}{\sigma^2}\right)^{\frac{\nu_0}{2}+1}\exp\left(-\frac{\nu_0\lambda_0}{2\sigma^2}\right) \sim \chi^{-2}(\nu_0,\lambda_0) \tag{4.6}$$

   or

$$\frac{\nu_0\lambda_0}{\sigma^2} \sim \chi^2(\nu_0). \tag{4.7}$$

4. Assuming the binary PPM symbols $x_k(i)$'s being independent, with $k \in \{1, 2, ..., K\}$ and i being the corresponding symbol of the kth user, the prior distribution $p(X)$ can be expressed as

$$p(X) = \prod_{i=0}^{M-1}\prod_{k=0}^{K} p_k(i)^{x_k(i)}[1-p_k(i)]^{1-x_k(i)} \tag{4.8}$$

where $p_k(i) = P(x_k(i) = 1)$.

4.1.2. Conditional Posterior Distributions

The following conditional distributions are needed for Gibbs sampler estimation steps with their derivations also found in Appendix A, derived in [17]:

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

$$p(\mathbf{a}|\sigma^2, \mathbf{X}, \mathbf{Y}) \propto N(\mathbf{a}_*, \mathbf{\Sigma}_*) I_{\mathbf{a}>0} \tag{4.9}$$

with

$$\mathbf{a}_* = \mathbf{\Sigma}_* \left(\mathbf{\Sigma}_0^{-1} \mathbf{a}_0 + \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \mathbf{B}(i) \mathbf{H}^T \mathbf{y}(i) \right) \tag{4.10}$$

$$\mathbf{\Sigma}_*^{-1} = \mathbf{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \mathbf{B}(i) \mathbf{R} \mathbf{B}(i), \tag{4.11}$$

where $\mathbf{R} = \mathbf{H}^T \mathbf{H}$, and $\mathbf{a_0}$ and $\mathbf{\Sigma}_0$ are defined in (4.5).

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

$$p(\sigma^2|\mathbf{a}, \boldsymbol{X}, \mathbf{Y}) \propto \chi^{-2} \left(\nu_0 + KM, \frac{\nu_0 \lambda_0 + s^2}{\nu_0 + KM} \right) \tag{4.12}$$

with

$$s^2 = \sum_{i=0}^{M-1} \|\mathbf{y}(i) - \mathbf{H}\mathbf{A}\mathbf{x}(i)\|^2 \tag{4.13}$$

where $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{x}$ are defined in chapter 2.

3. The conditional probabilities of $x_k(i) = 0$ or 1, given $\mathbf{a}, \sigma^2, \mathbf{X}_{ki}$ and $\mathbf{Y}$ with $\mathbf{X}_{ki}$

defined in part 4.1.3 can be obtained from

$$\frac{P[x_k(i)=1|\mathbf{a},\sigma^2,\boldsymbol{X}_{ki},\mathbf{Y}]}{P[x_k(i)=0|\mathbf{a},\sigma^2,\boldsymbol{X}_{ki},\mathbf{Y}]}=\frac{p_k(i)}{1-p_k(i)}\exp\left(\frac{2A_k}{\sigma^2}\mathbf{h}_k^T\left[\mathbf{y}(i)-\mathbf{H}\mathbf{A}\mathbf{x}_k^0(i)\right]\right) \quad (4.14)$$

with $p_k(i)$ being the *a priori* probability of a transmitted symbol 1 and

$$\mathbf{x}_k^0(i)=[x_1(i),...,x_{k-1}(i),0,x_{k+1}(i),...,x_K(i)]^T. \quad (4.15)$$

4.1.3. Gibbs Multiuser Detector

Our Gibbs sampler, given the initial values $\theta^{(0)}=[\mathbf{a}^{(0)}\ \sigma^{2(0)}\ \boldsymbol{X}^{(0)}]^T$ and posterior distributions, iterates the following loop:

- *Draw sample* $\boldsymbol{a}^{(n)}$ *from* $p(\boldsymbol{a}|\sigma^{2(n-1)},\mathbf{X}^{(n-1)},\boldsymbol{Y})$ *given by* (4.9).
- *Draw sample* $\sigma^{2(n)}$ *from* $p(\sigma^2|\boldsymbol{a}^{(n)},\mathbf{X}^{(n-1)},\boldsymbol{Y})$ *given by* (4.12).
- *For* i = 0, 1, ... , M-1

  For $k=1,2,...,K$

  Draw sample $x_k(i)^{(n)}$ *from* $P(x_k(i)|\boldsymbol{a}^{(n)},\sigma^{2(n)},\boldsymbol{X}_{ki}^{(n)},\boldsymbol{Y})$ *given by* (4.14) *where*
  $\boldsymbol{X}_{ki}^{(n)}=\{\boldsymbol{x}(0)^{(n)},...,\boldsymbol{x}(i-1)^{(n)},x_1(i)^{(n)},...,x_{k-1}(i)^{(n)},x_{k+1}(i)^{(n-1)},...,$
  $x_K(i)^{(n-1)},\boldsymbol{x}(i+1)^{(n-1)},...,\boldsymbol{x}(M-1)^{(n-1)}\}$.

The a posteriori symbol probabilities are approximated as

$$P(x_k(i)=1|\mathbf{Y})=\frac{1}{N}\sum_{n=n_0+1}^{n=n_0+N}\delta_{ki}^{(n)} \quad (4.16)$$

where n_0 is the "burn-in period", $\delta_{ki}^{(n)}=1$ if $x_k^{(n)}=1$ and $\delta_{ki}^{(n)}=0$ if $x_k^{(n)}=0$. A maximum *a posteriori* probability (MAP) decision is given by

$$\widehat{x}_k(i)=\arg\max_{b\in\{1,0\}}P[x_k(i)=b|\mathbf{Y}]. \quad (4.17)$$

4.2. Multiuser Detection in Impulsive Noise

In many realistic communication channels, the ambient noise is known to be impulsive. In this part, the Gibbs multiuser detector for an UWB system in impulsive noise is developed. The noise pdf and other parameters are defined in part 4.2.1. in detail. If we define an indicator random variable $I_j(i)$

$$I_j(i) = \begin{cases} 1 & \text{if} \quad n_j(i) \sim N(0, \sigma_1^2) \\ 2 & \text{if} \quad n_j(i) \sim N(0, \sigma_2^2) \end{cases} \tag{4.18}$$

for *i = 0, ..., M-1*; *j = 0, ..., N_h-1* with N_h being the number of allowed time hopping bins, given in chapter 2, then we shall denote $\mathbf{I} = \{I_j(i)\}_{j=0;\, i=0}^{N_h-1;\, M-1}$. We can define

$$\mathbf{\Lambda}(i) = \text{diag}\left(\sigma^2_{I_0(i)}, \sigma^2_{I_1(i)}, ..., \sigma^2_{I_{N_h-1}(i)}\right) \tag{4.19}$$

with *i = 0, ..., M-1*. Our unknown quantity vector will be defined as $\theta = [\mathbf{a} \;\; \sigma_1^2 \;\; \sigma_2^2 \;\; \epsilon \;\; \mathbf{I} \;\; \mathbf{X}]^T$. Then,

$$p(\theta|\mathbf{Y}) \sim P(\mathbf{Y}|\mathbf{a}, \sigma_1^2, \sigma_2^2, \epsilon, \mathbf{I}, \mathbf{X}) p(\mathbf{a})\; p(\sigma_1^2)\; p(\sigma_2^2)\; p(\epsilon)\; p(\mathbf{I}|\epsilon)\; p(\mathbf{X}) \tag{4.20}$$

$$\sim \quad \exp\left(-\frac{1}{2}\sum_{i=0}^{M-1}[\mathbf{y}(i) - \mathbf{HAx}(i)]^T \mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i) - \mathbf{HAx}(i)]\right)$$

$$\cdot \left(\frac{1}{\sigma_1^2}\right)^{(\frac{1}{2})\sum_{i=0}^{M-1} n_1(i)} \left(\frac{1}{\sigma_2^2}\right)^{(\frac{1}{2})\sum_{i=0}^{M-1} n_2(i)} p(\mathbf{a}) p(\sigma_1^2) p(\sigma_2^2) p(\epsilon) p(\mathbf{I}|\epsilon) p(\mathbf{X}) \tag{4.21}$$

where $n_l(i)$ is the number of $l's$ in $\{I_0(i), I_1(i), ..., I_{N_h-1}(i)\}, l = 1, 2$. [Note that $n_1(i) + n_2(i) = N_h$.] We employ Gibbs sampler in order to avoid the direct evaluation of the Bayesian estimate by integrating out all the elements of vector θ except for $\mathbf{X}$.

4.2.1. Prior Distributions

The following are the basic assumptions due to considerations of [17]:

1. The ambient noise distribution is a common two-term Gaussian mixture. The pdf

of n(i), serving as an approximation to the more fundamental Middleton Class A noise model [43], is given by

$$p(n_j(i)) = \frac{1-\epsilon}{(2\pi\sigma_1^2)^{1/2}} \exp\left(-\frac{n_j(i)^2}{2\sigma_1^2}\right) + \frac{\epsilon}{(2\pi\sigma_2^2)^{1/2}} \exp\left(-\frac{n_j(i)^2}{2\sigma_2^2}\right) \tag{4.22}$$

where $j = 0, ..., N_h - 1$, $0 < \epsilon < 1$ and $\sigma_1^2 < \sigma_2^2$.

2. For the unknown amplitude vector $\mathbf{a} = [A_1 \quad A_2 \quad ... \quad A_K]^T$ where A_k's are the channel coefficients, a truncated Gaussian prior distribution is assumed with

$$p(\mathbf{a}) \propto N(\mathbf{a}_0, \mathbf{\Sigma}_0) I_{a>0} \tag{4.23}$$

where $I_{a>0}$ is an indicator that is 1 if all the elements of $\mathbf{a}$ are positive and is 0 otherwise. This means, the algorithm can be used for positive channel amplitudes only.

3. For the noise variances σ_l^2, l = 1,2, independent inverse chi-square prior distributions are assumed

$$p(\sigma_l^2) \sim \chi^{-2}(\nu_l, \lambda_l), \quad l = 1, 2, \text{ with } \nu_1\lambda_1 < \nu_2\lambda_2. \tag{4.24}$$

4. Assuming the binary PPM symbols $x_k(i)$'s being independent, with $k \in \{1, 2, ..., K\}$ and i being the corresponding symbol of the kth user, the prior distribution $p(X)$ can be expressed as

$$p(X) = \prod_{i=0}^{M-1} \prod_{k=0}^{K} p_k(i)^{x_k(i)} [1 - p_k(i)]^{1-x_k(i)} \tag{4.25}$$

where $p_k(i) = P\left(x_k(i) = 1\right)$.

5. For the impulse probability ϵ, a β (denoting Beta distribution) prior distribution is assumed given by

$$p(\epsilon) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \epsilon^{a_0 - 1} (1 - \epsilon)^{b_0 - 1} \sim \beta(a_0, b_0). \tag{4.26}$$

6. Given impulse probability ϵ, the conditional distribution of the indicator random variable $I_j(i)$ is

$$\begin{aligned} p(I_j(i) = 1|\epsilon) &= 1 - \epsilon, & (4.27) \\ p(I_j(i) = 2|\epsilon) &= \epsilon & (4.28) \\ p(\mathbf{I}|\epsilon) &= (1-\epsilon)^{m_1}\epsilon^{m_2} & (4.29) \end{aligned}$$

with

$$\begin{aligned} m_1 &= \sum_{i=0}^{M-1} n_1(i) \\ m_2 &= \sum_{i=0}^{M-1} n_2(i) = MN_h - m_1. \end{aligned}$$

4.2.2. Conditional Posterior Distributions

The following conditional distributions are needed for Gibbs sampler estimation steps with their derivations found in Appendix B, derived in [17]:

1. The conditional distribution of the amplitude vector $\mathbf{a}$, given σ_1^2, σ_2^2, ϵ, $\mathbf{I}$, $\mathbf{X}$ and $\mathbf{Y}$ is given by

$$p(\mathbf{a}|\sigma_1^2, \sigma_2^2, \epsilon, (I), \mathbf{X}, \mathbf{Y}) \propto N(\mathbf{a}_*, \mathbf{\Sigma}_*) I_{\mathbf{a}>0} \quad (4.30)$$

with

$$\begin{aligned} \mathbf{a}_* &= \mathbf{\Sigma}_* \left(\mathbf{\Sigma}_0^{-1}\mathbf{a}_0 + \sum_{i=0}^{M-1} \mathbf{B}(i)\mathbf{H}^T\mathbf{\Lambda}(i)^{-1}\mathbf{y}(i) \right) & (4.31) \\ \mathbf{\Sigma}_*^{-1} &= \mathbf{\Sigma}_0^{-1} + \sum_{i=0}^{M-1} \mathbf{B}(i)\mathbf{H}^T\mathbf{\Lambda}(i)^{-1}\mathbf{H}\mathbf{B}(i). & (4.32) \end{aligned}$$

2. The conditional distribution of the noise variance σ_l^2 given $\mathbf{a}$, $\sigma_{\bar{l}}^2$, ϵ, $\mathbf{I}$, $\mathbf{X}$ and $\mathbf{Y}$

for $\bar{l} = 2$ if $l = 1$ and $\bar{l} = 1$ if $l = 2$, is given by

$$p(\sigma_l^2|\mathbf{a}, \sigma_{\bar{l}}^2, \epsilon, \mathbf{I}, \mathbf{X}, \mathbf{Y}) \quad \propto \quad \chi^{-2}\left([\nu_l + \sum_{i=0}^{M-1} n_l(i)], \frac{\nu_l\lambda_l + s_l^2}{\nu_l + \sum_{i=0}^{M-1} n_l(i)}\right) \quad (4.33)$$

with

$$s_l^2 = \sum_{i=0}^{M-1}\sum_{j=0}^{N_h-1} [y_j(i) - \xi_j^T \mathbf{A}\mathbf{x}(i)]^2 . 1_{\{I_j(i)=l\}} \quad (4.34)$$

where $1_{\{I_j(i)=l\}}$ is the indicator function such that it is 1 if $I_j(i) = l$, and it is 0 if $I_j(i) \neq l$; ξ_j^T is the *jth* row of the spreading waveform matrix $\mathbf{H}$, $j = 0, ..., N_h - 1$.

3. The conditional probability of $x_k(i) = 0$ or 1, given $\mathbf{a}$, σ_1^2, σ_2^2, ϵ, $\mathbf{I}$, $\mathbf{X}_{ki}$ and $\mathbf{Y}$ with $\mathbf{X}_{ki}$ denoting the set containing all elements of $\mathbf{X}$ except for $x_k(i)$, can be obtained from

$$\frac{P[x_k(i) = 1|\mathbf{a}, \sigma_1^2, \sigma_2^2, \epsilon, \mathbf{I}, \mathbf{X}_{ki}, \mathbf{Y}]}{P[x_k(i) = 0|\mathbf{a}, \sigma_1^2, \sigma_2^2, \epsilon, \mathbf{I}, \mathbf{X}_{ki}, \mathbf{Y}]} = \frac{p_k(i)}{1 - p_k(i)} \exp\left(2A_k \mathbf{h}_k^T \mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i) - \mathbf{H}\mathbf{A}\mathbf{x}_k^0(i)]\right) \quad (4.35)$$

with $p_k(i)$ being the *a priori* probability of a transmitted symbol 1, *k = 1, ..., K*; *i = 0, ..., M-1*, and

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

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

$$\frac{P[I_j(i) = 1|\mathbf{a}, \sigma_1^2, \sigma_2^2, \epsilon, \mathbf{I}_{ji}, \mathbf{X}, \mathbf{Y}]}{P[I_j(i) = 2|\mathbf{a}, \sigma_1^2, \sigma_2^2, \epsilon, \mathbf{I}_{ji}, \mathbf{X}, \mathbf{Y}]} = \frac{1-\epsilon}{\epsilon}(\frac{\sigma_2^2}{\sigma_1^2})^{1/2} \exp\left(\frac{1}{2}[y_j(i) - \xi_j^T \mathbf{A}\mathbf{x}(i)]^2(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2})\right) \quad (4.36)$$

for $j = 0, ..., N_h - 1$ and *i = 0, ..., M-1*.

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

$$P[\epsilon|\mathbf{a}, \sigma_1^2, \sigma_2^2, \mathbf{I}, \mathbf{X}, \mathbf{Y}] = \beta\left(a_0 + \sum_{i=0}^{M-1} n_2(i), b_0 + \sum_{i=0}^{M-1} n_1(i)\right). \quad (4.37)$$

4.2.3. Gibbs Multiuser Detector

Our Gibbs sampler, given the initial values $\theta^{(0)} = [\mathbf{a}^{(0)}\ \sigma_1^{2(0)}\ \sigma_2^{2(0)}\ \epsilon^{(0)}\ \mathbf{I}^{(0)}\ \mathbf{X}^{(0)}]$ and posterior distributions, iterates the following loop:

- *Draw sample* $\boldsymbol{a}^{(n)}$ *from* $p(\boldsymbol{a}|\sigma_1^{2(n-1)}, \sigma_2^{2(n-1)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ *given by* (4.30).
- *Draw sample* $\sigma_1^{2(n)}$ *from* $p(\sigma_1^2|\boldsymbol{a}^{(n)}, \sigma_2^{2(n-1)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ *given by* (4.33)*;*
  Draw sample $\sigma_2^{2(n)}$ *from* $p(\sigma_2^2|\boldsymbol{a}^{(n)}, \sigma_1^{2(n)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}^{(n-1)}, \boldsymbol{Y})$ *given by* (4.33).
- *For* i = 0, 1, ... , M-1

  For k = 1, 2, ..., K

  Draw sample $x_k(i)^{(n)}$ *from* $P(x_k(i)|\boldsymbol{a}^{(n)}, \sigma_1^{2(n)}, \sigma_2^{2(n)}, \epsilon^{(n-1)}, \boldsymbol{I}^{(n-1)}, \boldsymbol{X}_{ki}^{(n)}, \boldsymbol{Y})$ *given by* (4.35)*, where*
  $\boldsymbol{X}_{ki}^{(n)} = \{\boldsymbol{x}(0)^{(n)}, ..., \boldsymbol{x}(i-1)^{(n)}, x_1(i)^{(n)}, ..., x_{k-1}(i)^{(n)}, x_{k+1}(i)^{(n-1)}, ...,$
  $x_K(i)^{(n-1)}, \boldsymbol{x}(i+1)^{(n-1)}, ..., \boldsymbol{x}(M-1)^{(n-1)}\}$.
- *For* i = 0, 1, ... , M-1

  For $j = 1, 2, ..., N_h - 1$

  Draw $I_j(i)^{(n)}$ *from* $P(I_j(i)|\boldsymbol{a}^{(n)}, \sigma_1^{2(n)}, \sigma_2^{2(n)}, \epsilon^{(n-1)}, \boldsymbol{I}_{ji}^{(n)}, \boldsymbol{X}^{(n)}, \boldsymbol{Y})$ *given by* (4.36)*, where*
  $\boldsymbol{I}_{ji}^{(n)} = \{\boldsymbol{I}_0(0)^{(n)}, ..., \boldsymbol{I}_{N_h-1}(i-1)^{(n)}, I_0(i)^{(n)}, ..., I_{j-1}(i)^{(n)}, I_{j+1}(i)^{(n-1)}, ...,$
  $I_{N_h-1}(i)^{(n-1)}, \boldsymbol{I}_{N_h-1}(M-1)^{(n-1)}\}$.
- *Draw* $\epsilon^{(n)}$ *from* $p(\epsilon|\boldsymbol{a}^{(n)}, \sigma_1^{2(n)}, \sigma_2^{2(n)}, \boldsymbol{I}^{(n)}, \boldsymbol{X}^{(n)}, \boldsymbol{Y})$ *given by* (4.37).

The a posteriori symbol probabilities are approximated as

$$P(x_k(i) = 1|\mathbf{Y}) = \frac{1}{N} \sum_{n=n_0+1}^{n=n_0+N} \delta_{ki}^{(n)} \tag{4.38}$$

where n_0 is the "burn-in period", $\delta_{ki}^{(n)} = 1$ if $x_k^{(n)} = 1$ and $\delta_{ki}^{(n)} = 0$ if $x_k^{(n)} = 0$. A MAP decision is given by

$$\widehat{x}_k(i) = \arg \max_{b \in \{1,0\}} P[x_k(i) = b|\mathbf{Y}]. \tag{4.39}$$

4.3. Multiuser Detection in Cauchy Noise

In many different fields such as telecommunications, physics and finance, Gaussian and even impulsive models reveal difficulties in fitting data that exhibits a high degree of heterogeneity; thus stable distributions have been introduced. Alpha stable noise is a generalization of the Gaussian model [44]. Stable distributions allow also for infinite variance, skewness and heavy tails. A summary of stable distributions can be found in [48] and [49], which provide a good theoretical background on heavy-tailed distributions. The practical use of heavy-tailed distributions in many different fields is well documented in [50], which also reviews the estimation techniques. First, a definition of the stable distributions need to be made. A random variable X has stable distribution $S_\alpha(\beta, \mathrm{t}, \mathrm{s})$ if its parameters are in the following ranges: $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\mathrm{t} \in (-\infty, +\infty)$, $\mathrm{s} \in (0, +\infty)$ and if its characteristic function can be written as

$$E[\exp(i\vartheta x)] = \begin{cases} \exp\left(-|\mathrm{s}\vartheta|^\alpha\right)\left(1 - i\beta(\mathrm{sign}(\vartheta))\tan(\pi\alpha/2) + i\mathrm{t}\vartheta\right) & \text{if } \alpha \neq 1 \\ \exp\left(-|\mathrm{s}\vartheta|\left(1 + 2i\beta\ln|\vartheta|\mathrm{sign}(\vartheta)/\pi\right) + i\mathrm{t}\vartheta\right) & \text{if } \alpha = 1 \end{cases} \tag{4.40}$$

where ϑ is a real number. The stable distribution is thus completely characterized through the following four parameters: the characteristic exponent α, the skewness parameter β, the location parameter t and finally the scale parameter s.

Stable distributions admit explicit representation of the density function only in the following cases: the Gaussian distribution $S_2(0, \mathrm{s}, \mathrm{t})$, the Cauchy distribution $S_1(0, \mathrm{s}, \mathrm{t})$ and the Levy distribution $S_{1/2}(1, \mathrm{s}, \mathrm{t})$ [44, 45]. As our aim is not to approximate the probability density function, but to obtain the system performance under such noise, our model makes use of the Cauchy distribution as it has a closed form expression.

In this part, the Gibbs multiuser detector is developed for a two stage scenario. The Gibbs sampler is held the same as for the impulsive noise scenario given in detail in the previous section. At the first stage, the sampler is run for an estimation of the channel coefficients. After this first stage and depending on the estimated channel co-

efficients, a second run is applied for estimating the symbol values. This was required, because sampling from a multivariate and unconventional distribution, such as the conditional posterior distribution of the channel coefficient vector $\mathbf{a}$, is another research area with no obvious and straightforward solution [46, 47]. Also, sampling from a univariate and unconventional distribution is a challenging issue to resolve. The rejection sampling solution is applied as a solution, about which, detailed information can be found in [24, 47]. The structure for this scenario is given in detail in the following subsections. However, this suboptimal approach did not work out, as the convergence of the Gibbs sampler is very hard to maintain. But, certain work has been carried out, thus a detailed description for this model will also be provided for the sake of future studies.

4.3.1. Prior Distributions

The following are the basic assumptions for the prior distributions:

1. The ambient noise distribution is a Cauchy distribution. The pdf is given by

$$p(n(i)) = \frac{s}{\pi\left(s^2 + (n(i) - t)^2\right)} \tag{4.41}$$

.

2. For the unknown amplitude vector $\mathbf{a} = [A_1 \quad A_2 \quad ... \quad A_K]^T$ where A_k's are the channel coefficients, a truncated Gaussian prior distribution is assumed with

$$p(\mathbf{a}) = N(\mathbf{a}_0, \mathbf{\Sigma}_0)_{I_{\{\mathbf{a}>0\}}} \tag{4.42}$$

where $I_{a>0}$ is an indicator that is 1 if all the elements of $\mathbf{a}$ are positive and is 0 otherwise.

3. For the location parameter t, a Gaussian prior distribution is assumed

$$p(t) = N(\mu_{\mathbf{t}}, \mathbf{\Sigma_t}). \tag{4.43}$$

4. For the scale parameter s, an inverse chi-square prior distribution is assumed

$$\begin{aligned} p(s^2) &= \frac{\left(\frac{\nu_0\lambda_0}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)}\left(\frac{1}{s^2}\right)^{\frac{\nu_0}{2}+1}\exp\left(-\frac{\nu_0\lambda_0}{2s^2}\right) \\ &\sim \chi^{-2}(\nu_0,\lambda_0) \end{aligned} \tag{4.44}$$

5. Assuming the binary PPM symbols $x_k(i)$'s being independent, with $k \in \{1, 2, ..., K\}$ and i being the corresponding symbol of the kth user, the prior distribution $p(X)$ can be expressed as

$$p(\mathbf{X}) = \prod_{i=0}^{M-1}\prod_{k=0}^{K} p_k(i)^{x_k(i)}[1-p_k(i)]^{1-x_k(i)} \tag{4.45}$$

where $p_k(i) = P(x_k(i) = 1)$.

4.3.2. Conditional Posterior Distributions

The following conditional distributions with their derivations found in Appendix C, are used in the simulations:

1. The conditional distribution of the amplitude vector **a**, given t, s, **X** and **Y** is given by

$$p(\mathbf{a}|t,s,\mathbf{X},\mathbf{Y}) = \frac{s}{\pi\left(s^2+[(\sum_{i=0}^{M-1} y(i) - \mathbf{HB(i)a} - t]^2\right)}\exp\left(-\frac{1}{2}(\mathbf{a}-\mathbf{a_0})\mathbf{\Sigma}_0^{-1}(\mathbf{a}-\mathbf{a_0})^T\right). \tag{4.46}$$

2. The conditional distribution of the location parameter t, given **a**, s, **X** and **Y** is given by

$$p(t|\mathbf{a},s,\mathbf{X},\mathbf{Y}) = \frac{s}{\pi\left(s^2+[(\sum_{i=0}^{M-1} y(i) - \mathbf{HA}x(i)) - t]^2\right)}\exp\left(-\frac{1}{2}(\mathbf{t}-\mu_\mathbf{t})\mathbf{\Sigma}_t^{-1}(\mathbf{t}-\mu_\mathbf{t})^T\right). \tag{4.47}$$

3. The conditional distribution of the location parameter s, given **a**, t, **X** and **Y** is given by

$$p(s|\mathbf{a},t,\mathbf{X},\mathbf{Y}) = \frac{s}{\pi\left(s^2+[(\sum_{i=0}^{M-1} r(i) - \mathbf{HA}x(i)) - t]^2\right)}\left(\frac{1}{s}\right)^{1+(\nu_0/2)}\exp\left(-\frac{\nu_0\lambda_0}{2s}\right). \tag{4.48}$$

4. The conditional probability of $x_k(i) = 0$ or 1, given **a**, s, t, $\mathbf{X}_{ki}$ and **Y** with $\mathbf{X}_{ki}$ denoting the set containing all elements of **X** except for $x_k(i)$, can be obtained from

$$\frac{p(x_k(i)=1|s,\mathbf{a},t,\mathbf{X_{ki}},\mathbf{Y})}{p(x_k(i)=0|s,\mathbf{a},t,\mathbf{X_{ki}},\mathbf{Y})} = \frac{p_k(i)}{1-p_k(i)} \frac{\left(s^2 + [(\sum_{i=0}^{M-1} r(i) - \mathbf{HA}(x_k^0(i) + \mathbf{1}_k)) - t]^2\right)}{\left(s^2 + [(\sum_{i=0}^{M-1} y(i) - \mathbf{HA}(x_k^0(i) - \mathbf{1}_k)) - t]^2\right)} \tag{4.49}$$

with $p_k(i)$ being the *a priori* probability of a transmitted symbol 1, *k = 1, ..., K*; *i = 0, ..., M-1*, and

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

4.3.3. Gibbs Multiuser Detector

Our Gibbs sampler, given the initial values $\theta^{(0)} = [\mathbf{a}^{(0)} \ s^{(0)} \ t^{(0)} \ \mathbf{X}^{(0)}]$ and posterior distributions, iterates the same loop as in part 4.2.3. for the first stage where the channel coefficients are estimated. In the second phase, (assuming the correctly estimated channel coefficients are in hand) Cauchy noise parameters and information bits are tried to be estimated. As sampling from a univariate unconventional density is also a hard issue, rejection sampling is employed. For this purpose, two hat functions need to be determined one of which will have values higher than (4.47) and the other than (4.48). The lognormal distribution multiplied with an appropriate scaling factor (0.65 for this case) is used as shown in Figure 4.1. The rejection sampling algorithm works as :

1. Generate a random variable f from uniform distribution in the range $[0, 1]$.
2. Denote scaled lognormal distribution with $g(\theta)$ and the desired conditional posterior density as $p(\theta|y)$. Then, if $f > \frac{p(\theta|y)}{g(\theta)}$, then accept θ drawn from $g(\theta)$ as a draw from $p(\theta|y)$. If not, return to the first step.

The failure of the proposed scenario is because of the deviations in the estimated values for the first stage. The estimated channel coefficients do not converge to the true values unaffected by how long the burn-in period is. This is caused by the sensitive

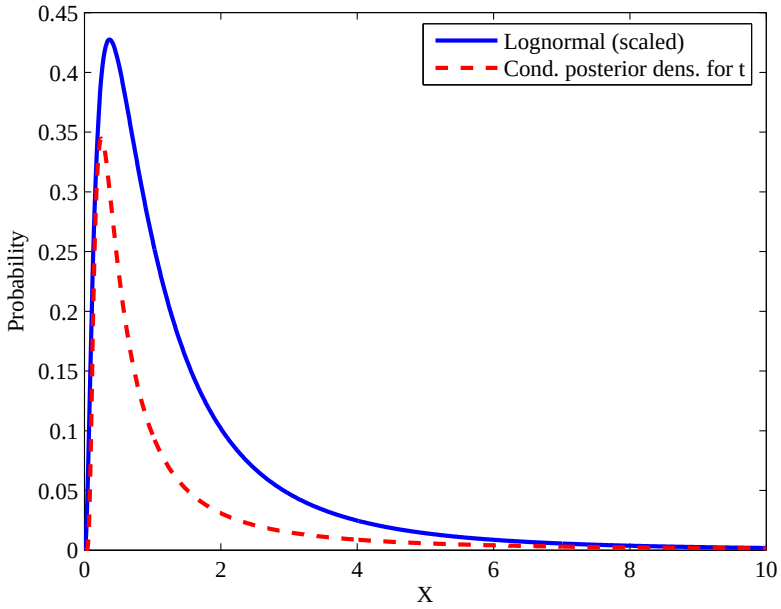

Figure 4.1. The hat function for rejection sampling.

convergence requirements of the Gibbs sampler. The sub-optimality of the scenario does not let the sampler converge. As the Gibbs sampler solution is not obtained, the results obtained in this part will not be discussed in the following sections except for the Conclusions part.

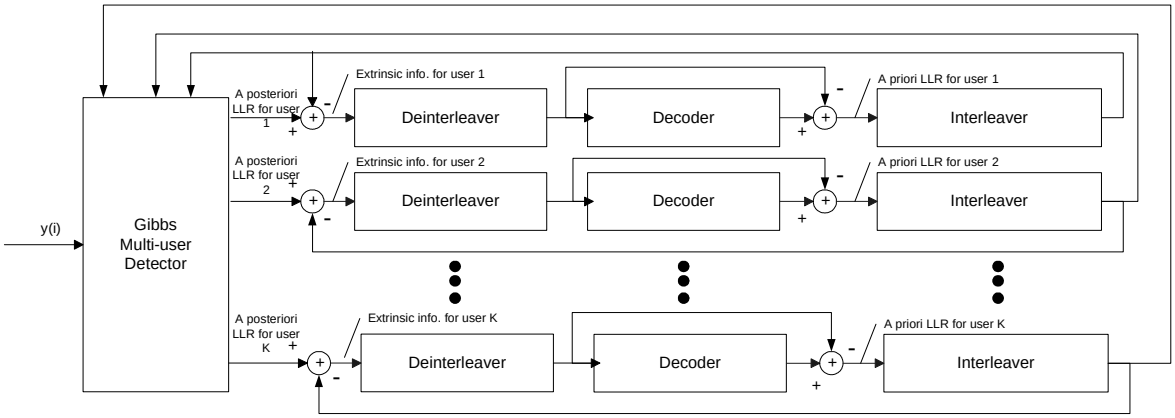

Figure 4.2. The iterative multiuser detector structure.

4.4. Adaptive Turbo Multiuser Detector

The Gibbs sampler makes hard decisions based on the *a posteriori* log-likelihood ratios (LLR) of the transmitted symbols given in (4.17) or (4.39) with respect to the noise type. However, its output contains the *a posteriori* conditional probabilities of $x_k(i)$ being 0 or 1 in (4.14) or (4.35) for Gaussian or impulsive noise respectively. These probabilities added and averaged over the after-burn-in steps of the Gibbs sampler can be used as the soft input to the iterative decoder. One can therefore obtain the *a posteriori* log-likelihood ratios depending on the symbol value of the transmitted symbol as

$$\lambda_1[x_k(i)] = \log \frac{P(x_k(i) = +1|\, \boldsymbol{Y})}{P(x_k(i) = 0|\, \boldsymbol{Y})} \tag{4.50}$$

for k = *1,2, ...,* K and i = *0, 1, ..., M-1.* The above equation can also be rewritten as

$$\lambda_1[x_k(i)] = \log \frac{p(\, \boldsymbol{Y}|x_k(i) = 1)}{p(\, \boldsymbol{Y}|x_k(i) = 0)} + \log \frac{P(x_k(i) = 1)}{P(x_k(i) = 0)}. \tag{4.51}$$

for k = *1,2, ...,* K and i = *0, 1, ..., M-1.* The second term of (4.52) is the *a priori* information of $x_k(i)$ computed by the channel decoder in the previous iteration, interleaved and fed back to the proposed multiuser detector for the following iteration. Let us denote this *a priori* information by $\varphi_2^p[x_k(i)]$, with the superscript indicating the quantity obtained from the previous iteration. It equals 0 for the first iteration as no prior information exists. The first term in (4.52) is the extrinsic information which is, after being deinterleaved, used by the channel decoder to produce *a posteriori* LLR's. Let us denote the first term by $\varphi_1[x_k(i)]$. As mentioned $\lambda_1[x_k(i)]$ is obtained by adding and averaging the conditional probabilities of a symbol being either 0 or 1 over converged iterations of the Gibbs sampler and it equals to the value of the first term for the first iteration of the decoder. These LLR's are after being interleaved, sent to the multiuser detector to be used in the calculation of the *a priori* distributions as shown in Figure 4.2. Note that for all the iterations, the decoder-produced *a priori* LLR's are interleaved and then subtracted from the multiuser detector output, the *a posteriori* LLR's, to form the extrinsic information not influenced by the a priori information

computed by the decoder in the previous iteration. The decoder in Figure 4.2 makes use of the well-known MAP algorithm [51]. If we name the output of the channel decoder as $\varsigma_2[x_k(m)]$ for $k = 1, 2, ..., K$ and $m = 0, 1, ..., M-1$, based on the extrinsic information of the code bits from the previous iteration $\{\varphi_1^p[x_k(i)]\}_{k=1;m=0}^{K;M-1}$ and the structure of the channel code, the soft-input soft-output channel decoder computes the *a posteriori* LLR of each code bit as

$$\varsigma_2[x_k(m)] = \log \frac{P\left[x_k(m) = 1 | \{\varphi_1^p[x_k(i)]\}_{k=1;i=0}^{K;M-1}; \text{code constraints}\right]}{P\left[x_k(m) = 0 | \{\varphi_1^p[x_k(i)]\}_{k=1;i=0}^{K;M-1}; \text{code constraints}\right]} = \varphi_2[x_k(m)] + \varphi_1^p[x_k(m)]. \tag{4.52}$$

It is seen from (4.53) that the output of the soft-input soft-output channel decoder is the sum of the prior information and the extrinsic information 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 $\{\varphi_1^p[x_k(l)]\}_{l \neq m}$ 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 $\{\varphi_2[x_k(m)]\}_{k=1;m=0}^{K;M-1}$ is then used to compute the *a priori* symbol distributions $\{p_k(i)\}_{k=1;i=0}^{K;M-1}$ defined in (4.8) from the corresponding LLR's as follows. Since we called

$$\varphi_2^p[x_k(i)] = \log \frac{P(x_k(i) = 1)}{P(x_k(i) = 0)} \tag{4.53}$$

after some manipulations as explained in detail in [17], we have

$$\begin{aligned} p_k(i) = P[x_k(i) = 1] &= \frac{\exp\left(\varphi_2^p[x_k(i)]\right)}{1 + \exp\left(\varphi_2^p[x_k(i)]\right)} \\ &= \frac{\exp\left(\frac{1}{2}\varphi_2^p[x_k(i)]\right)}{\exp\left(-\frac{1}{2}\varphi_2^p[x_k(i)]\right) + \exp\left(\frac{1}{2}\varphi_2^p[x_k(i)]\right)} \\ &= \frac{\cosh\left(\frac{1}{2}\varphi_2^p[x_k(i)]\right)\ \left[1 + \tanh\left(\frac{1}{2}\varphi_2^p[x_k(i)]\right)\right]}{2\cosh\left(\frac{1}{2}\varphi_2^p[x_k(i)]\right)} \\ &= \frac{1}{2}\ \left[1 + \tanh\left(\frac{1}{2}\varphi_2^p[x_k(i)]\right)\right]. \end{aligned} \tag{4.54}$$

The symbol probabilities $\{p_k(i)\}_{k=1;i=0}^{K;M-1}$ are then fed back to the Gibbs multiuser detector as the prior information for the next iteration. At the first iteration, the extrinsic information $\varphi_1[x_k(m)]$ and $\varphi_2[x_k(m)]$ are statistically independent. Subsequently, since they use the same information indirectly, they will become increasingly correlated, which causes the improvement through the iterations to diminish.

5. SIMULATION RESULTS

A binary PPM modulated UWB system of 114 users in a 114 coefficient non-line-of-sight UWB channel given in [38] with all the coefficients being positive is used in the simulations. 5 of the users are thought to be active in this system.

The system parameters are $N_f = 10$ frames / symbol, $N_h = 500$ bins, $N_c = 1000$ chips / frame. Time hopping was allowed in the first half of each frame in order to avoid inter-frame interference. The channel code for each user is a rate that is one half of the constraint length-5 convolutional code using the generator polynomial (23,35).

The initial values of $\mathbf{a}_0 = [1\ 1\ ...\ 1]^T$ being a vector of size 1 x 114, $\mathbf{\Sigma}_0 = 1000\boldsymbol{I}$ being a matrix of size 114 x 114 are used in (4.5) and (4.24), $\nu_0 = 1$, $\lambda_0 = 0.1$ are used in (4.6), $\nu_l = 1$, $\lambda_l = 0.1$ in (4.25) for $l = 1$ and $\nu_l = 1$, $\lambda_l = 1$ again in (4.25) for $l = 2$ in the simulations. Also $a_0 = 1$ and $b_0 = 2$ are used in (4.27). $\mathbf{H}$ is formed by trimming 128 bit long Walsh codes to 114 bits. $\epsilon = 0.1$ for the case of impulsive noise. Also $\kappa = 10$ for the same case.

In the simulations, different scenarios were tried for various values of ϵ and κ. It is observed that Gibbs sampler fails to converge for values of κ greater than 10. Also, $\epsilon = 0.5$ was tried, in which case, convergence problems arose one more time.

The simulation results are obtained using the Monte Carlo simulation method. The results for cases the Gibbs sampler does not converge to a solution are discarded.

The tracking performance of the Gibbs sampler in Gaussian noise is seen in Figure 5.1. As can be observed in the figure, the sampler converges to the true solution in about 10 - 12 iterations for this scenario. The estimated values around the true value after the burn-in period is limited to a very narrow range, which indicates a good estimation performance for the converged steps. The estimated values are as good at the first converged step as at any other converged step. The histograms of the same

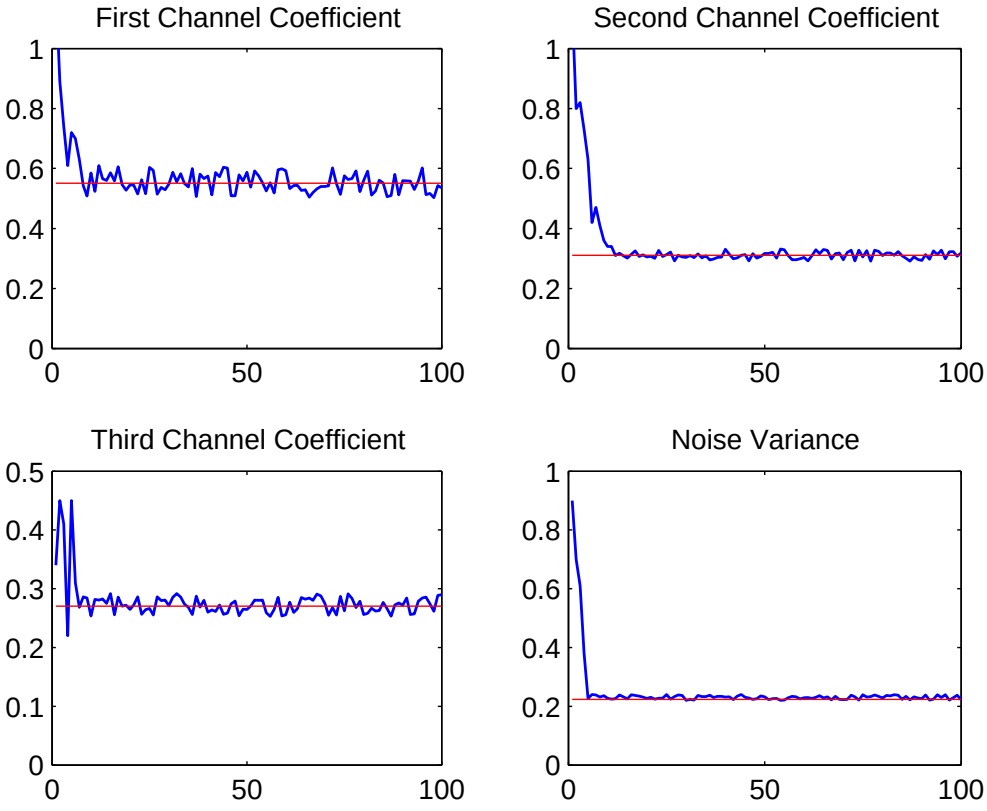

Figure 5.1. Gibbs sampler tracking performance in Gaussian noise at Eb/No = 4 dB. The actual values in order are [0.55095 0.3102 0.2702 0.22353].

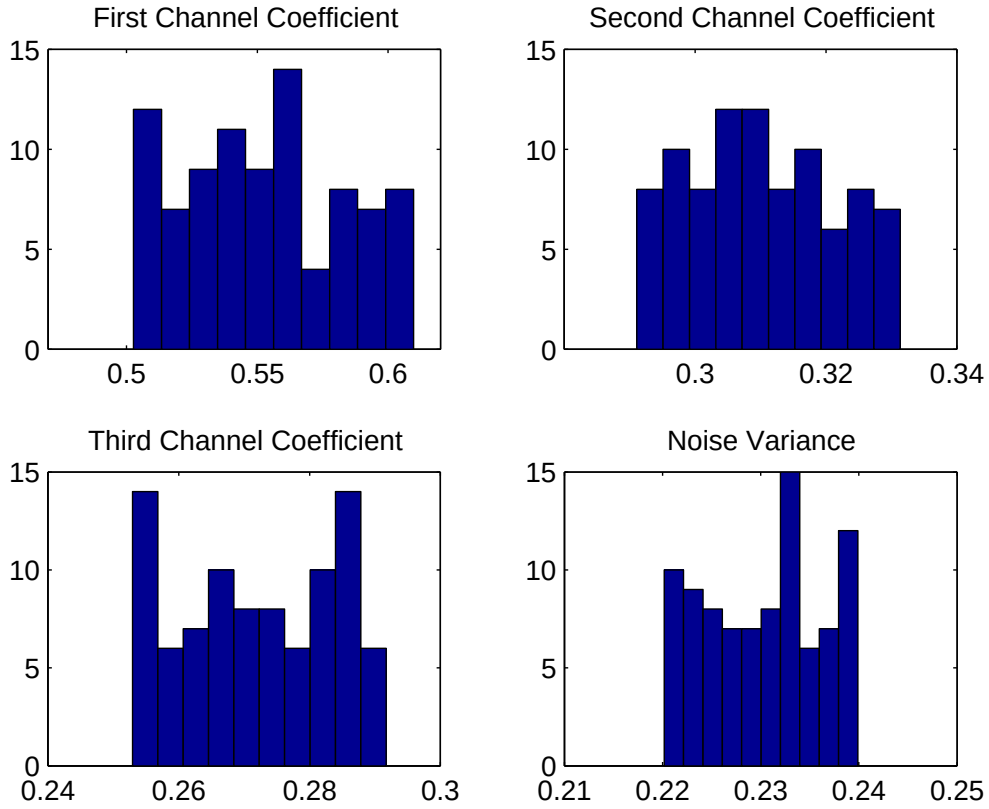

Figure 5.2. Gibbs sampler outputs for the 100 samples after the burn-in period of length 50 in Gaussian noise.

scenario is given by Figure 5.2. We can once more see that the estimated values are in a narrow range. The histograms are expected to have a much more concentrated shape around the true values as the number of converged iterations used to draw the histograms increases.

In Figure 5.3, we can see the performance of the correlator receiver as well as the Gibbs sampler performance for all 3 turbo iterations of interest. At a certain threshold, the first Gibbs sampler iteration starts to outperform the conventional receiver, which is assumed to have perfect channel information. Especially between the first and second turbo iterations, there is considerable gain. We can observe that third iteration adds very little performance improvement which puts forward a discussion on the gain between using two or three iterations. The sampler shall be evaluated according to the bit error rate (BER) requirements.

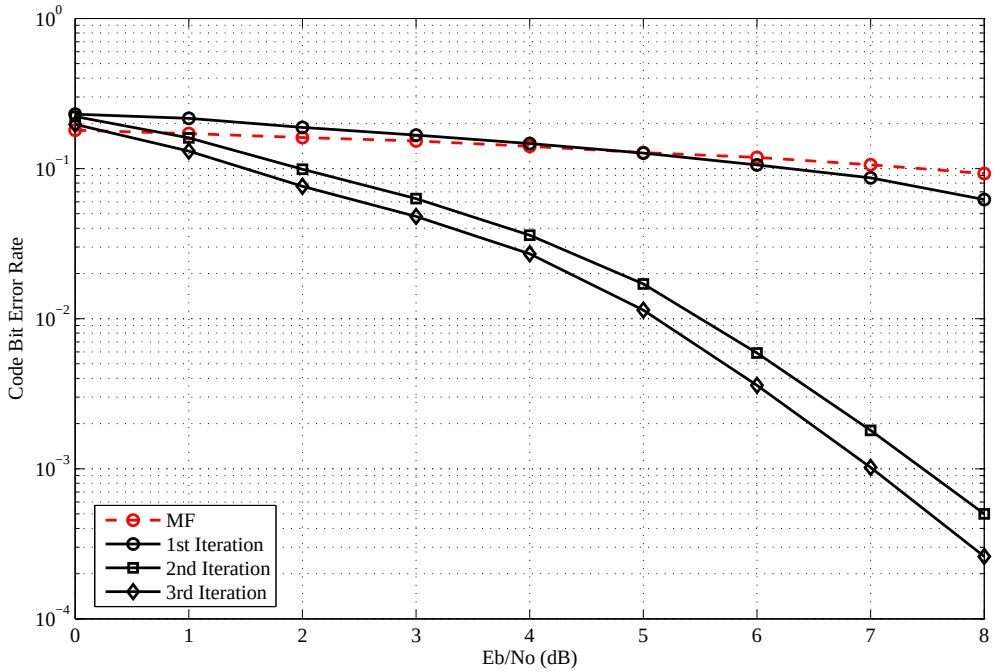

Figure 5.3. Bit error rate performance - convolutional code, Gaussian noise. Averaged over first 3 users. All users have the same amplitudes.

The tracking performance of the Gibbs sampler under impulsive noise distortion is given in Figure 5.4. As written in the caption, the noise variance value used in this simulation set-up is less than that is used in Figures 5.1, 5.2 and 5.3. This is because of the convergence requirements of the sampler. Due to this modification, a similar tracking performance of the Gibbs sampler is observed. The burn-in period is again about 10 - 15 steps. The histograms for the same scenario given in Figure 5.5 gives a more concentrated picture when compared to Figure 5.2. This is mostly because of the smaller noise variance value.

Figure 5.6 gives the comparative performance of the Gibbs sampler under Gaussian and impulsive noise, in the set-up used for obtaining the results in Figures 5.4 and 5.5. κ has the value 10 for the impulsive case and 1 for the Gaussian case. The impulse probability ϵ is set to be 0.1 for the impulsive scenario. The third iteration performance of the sampler under impulsive noise takes similar values of the second iteration for the Gaussian case, which means impulsiveness effects can be removed by one extra

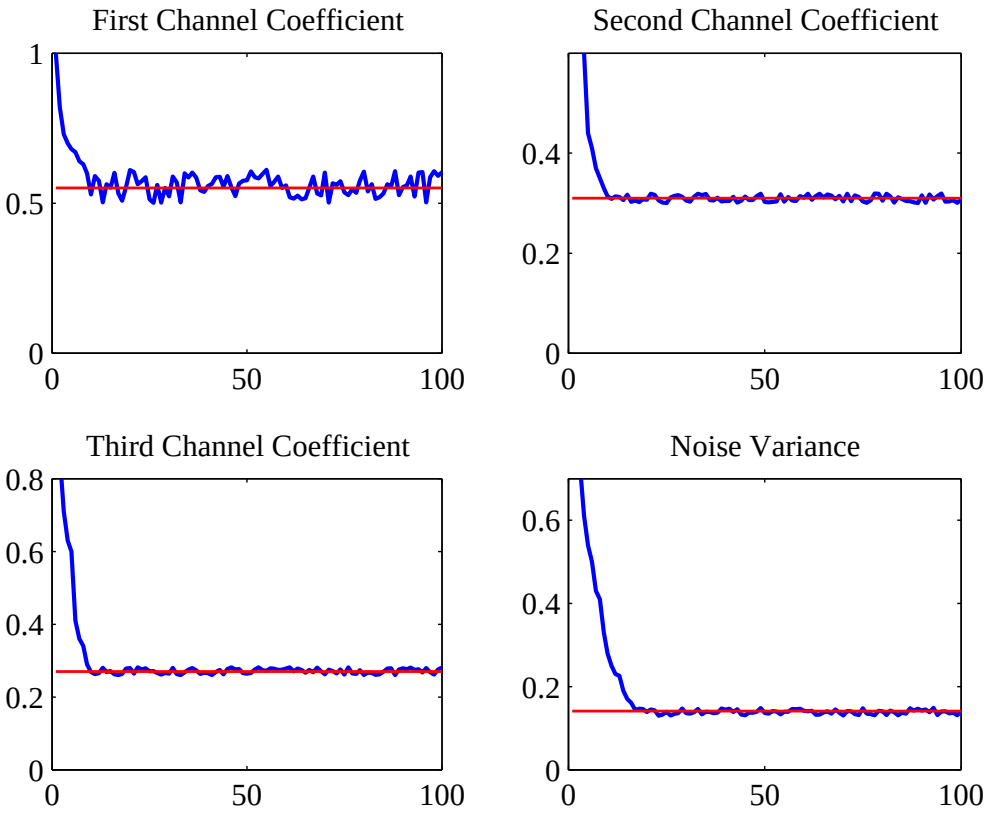

Figure 5.4. Gibbs sampler tracking performance for the 100 iterations after burn-in for the first 3 channel coefficients and noise variance.

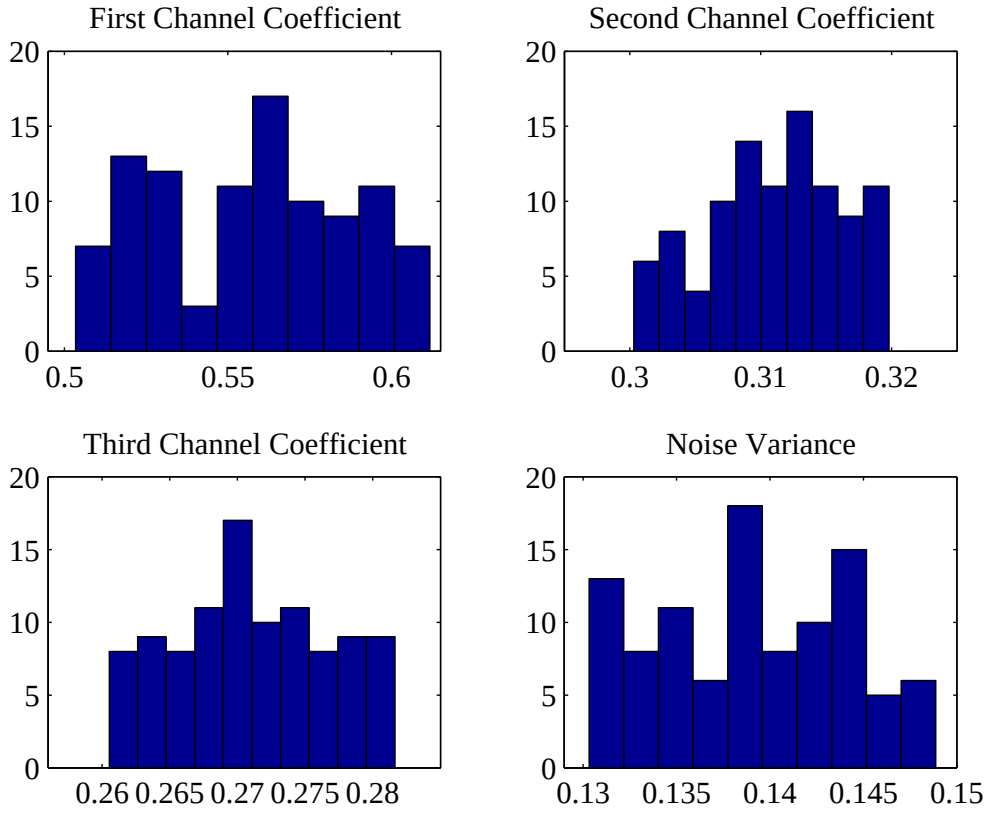

Figure 5.5. Gibbs sampler outputs for the 100 samples after the burn-in period of length 50 in impulsive noise. The actual values in order are [0.55095 0.3102 0.2702 0.141].

iteration. In the E_b/N_0 range given in Figure 5.6, the sampler gain for the Gaussian case between second and third iterations is more significant when compared to the same gain in Figure 5.3. The first iteration of the Gibbs sampler, which contains no iterative effect, achieves the same BER value as the third iteration at $E_b/N_0 = 28$ dB, which indicates that the iterative multiuser detector provides a gain of about 24 dB's.

The performance of the Gibbs sampler under impulsive noise is also given in comparison with an M-estimator based on the basic principles in [52] and [53] in Figure 5.7. The Gibbs sampler seems to perform better than the M-estimator, but the M-estimator is known to provide higher gains for higher κ values, where our sampler fails to converge. The M-estimator uses 32 fingers of the channel and is the robust version of the multipath-combining decorrelating detector in [53], which performs multipath

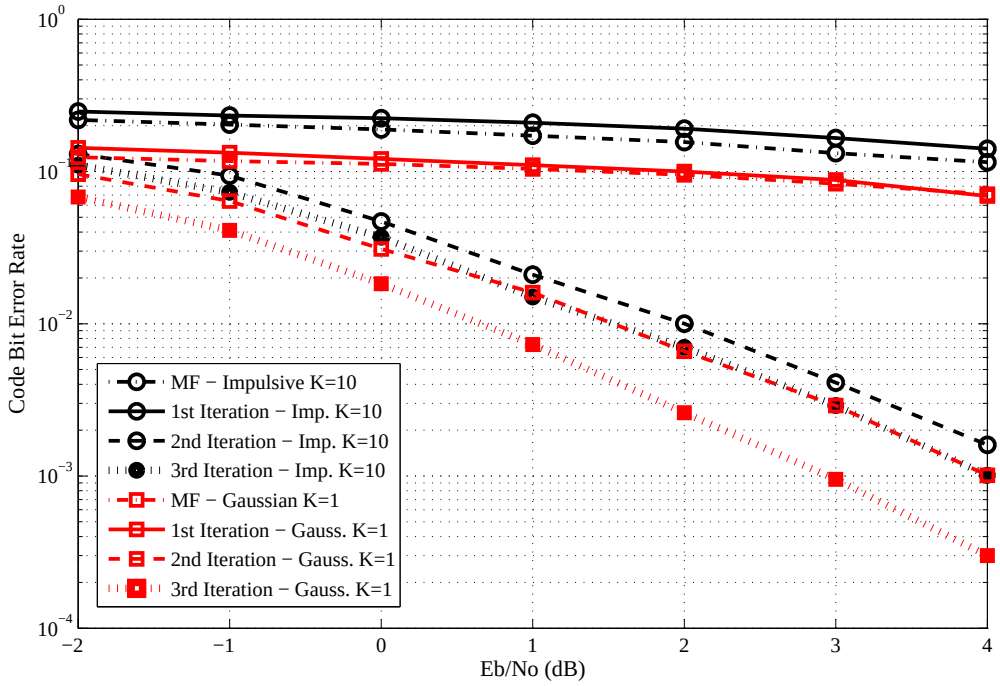

Figure 5.6. Bit error rate performance - convolutional code, Gaussian and Impulsive noise types. BER averaged over first 3 users. All users have the same amplitudes.

combining prior to decorrelation.

In general, we can say that the iterative multiuser detector proposed is superior to the conventional matched-filter receiver as the SNR is increased. But as the number of iterations is increased, the gain is getting smaller, as observed in all turbo systems. In Figures 5.3 and 5.6, the curves corresponding to all iterations are those at the output of the blind multiuser detector, first iteration being the uncoded bit error rate.

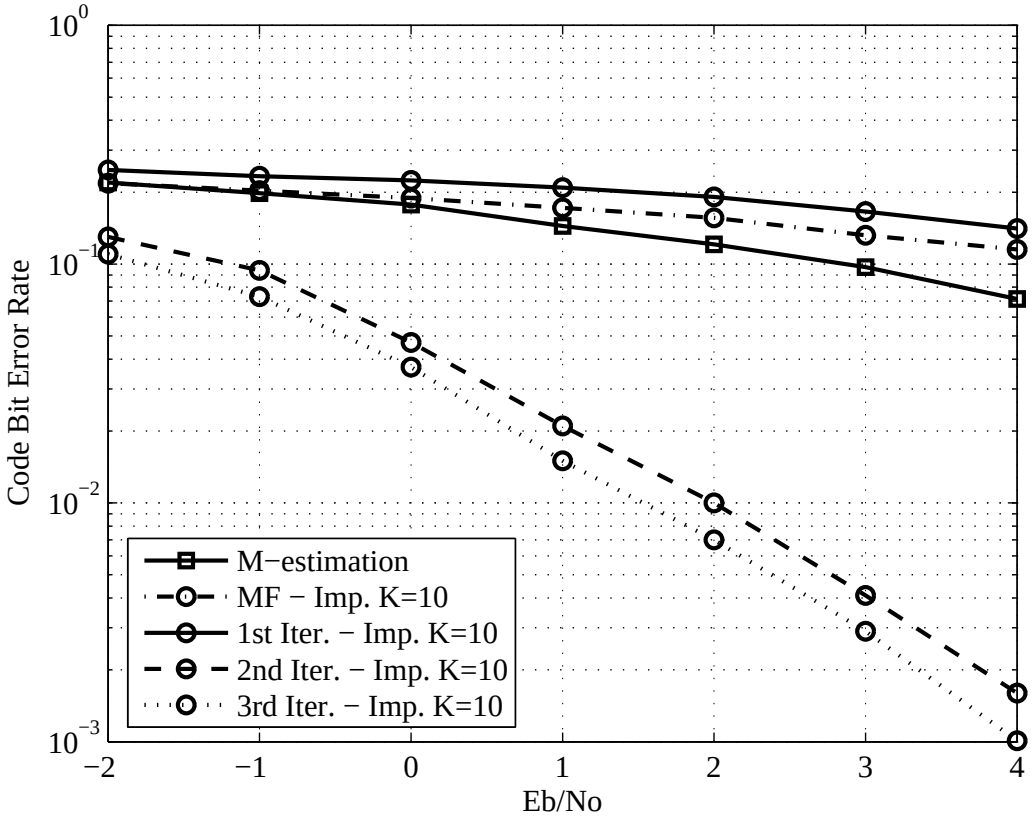

Figure 5.7. Bit error rate performance comparison of the Gibbs sampler and M-estimation.

6. CONCLUSIONS

In this thesis, a blind Bayesian multiuser detector is developed for an impulse radio UWB system. The Gibbs sampler receiver is adapted to a binary PPM modulated impulse radio UWB system in a frequency selective UWB channel. The soft outputs of the Gibbs sampler is iteratively decoded. Two scenarios were considered, with the difference between the two being the noise type : Gaussian and impulsive. This iterative multiuser detector is proven by the simulation results to be more effective than the conventional matched filter for UWB systems. Also the tracking performance of the sampler is assessed with the histograms and sampler outputs at each iteration.

The convergence of the sampler is not always guaranteed because of the UWB channel structure. That's why, for an UWB system, need on using boundaries arises while simulations. We can conclude that Gibbs sampler is not a miraculous solution. One other shortcoming is that the burn-in period for the sampler cannot be predicted, or theoretically calculated. It must also be added that Gibbs sampler outperforms the conventional receiver, but it also adds considerable complexity. The complexity of the solution is less than that of theoretical posterior density calculation, but far more than the conventional UWB receiver. The proposed system shall be evaluated with respect to the BER requirements of any design in which this system is going to be employed.

It must be concluded that especially when conjugate priors are used as in our Gaussian and impulsive scenarios, the conditional posterior distribution derivations and implementation are the easiest. As this kind of a situation results in the same form of posteriors, the calculation and random draws from the posteriors are easiest. Whenever a prior not conjugating the likelihood is used, due to the form of the posterior, the analytical expression requires more effort and random number generation problems, with a high probability, may arise.

As done in our model, UWB systems can be modeled by some means as a CDMA system and applications developed for CDMA can be applied also to UWB models.

Making small modifications such as replacing the inherent replication property by orthogonal coded spreading serves to this aim.

Frequency selective UWB channel brings in many difficulties as it has many taps and because of the signal model, the effects of the channel can be catastrophic. Because of this characteristic of the UWB channel, the implementation is more difficult than it would be for a shorter channel and convergence problems arose as mentioned. Also, the channel could cause inter-frame interference, which shall be taken into account while designing the allowable time hopping bins. This is taken into consideration in the selection of N_c and N_h values for our model.

Our solution in this thesis shall be evaluated both with its performance improvement capability and its complexity. Our system outperforms the conventional matched-filter receiver, but it certainly adds much complexity. It shall be considered where BER requirements are tight. The iterative decoding can even be done for only 2 iterations because the achieved BER level is quite satisfactory.

Many different applications can be further examined. The existence of new rejection methods for sampling from multivariate densities will allow the examination of the behavior of the system under alpha-stable noise, especially dealing with special and closed form expressible types, such as Cauchy or Levy distributions. The work in this thesis can be a starting point for such studies.

APPENDIX A: DERIVATION FOR THE GAUSSIAN CASE

Derivation of (4.9)

$$
\begin{aligned}
& p(\mathbf{a}|\sigma^2, \mathbf{X}, \mathbf{Y}) \\
= \;& p(\mathbf{a}, \sigma^2, \mathbf{X}|\mathbf{Y})/p(\sigma^2, \mathbf{X}|\mathbf{Y}) \\
\propto \;& p(\mathbf{a}, \sigma^2, \mathbf{X}|\mathbf{Y}) \qquad \text{(A.1)} \\
\propto \;& \exp[-\frac{1}{2\sigma^2}\sum_{k=0}^{M-1}\|\mathbf{y}(i) - \mathbf{HB}(i)\mathbf{a}\|^2]\exp[-\frac{1}{2}(\mathbf{a}-\mathbf{a}_0)^T\Sigma_0^{-1}(\mathbf{a}-\mathbf{a}_0)] \\
\propto \;& \exp\left\{-\frac{1}{2}\mathbf{a}^T[\mathbf{\Sigma_0^{-1}} + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\mathbf{B(i)H^THB(i)}]\mathbf{a} + \mathbf{a}^T[\mathbf{\Sigma_0^{-1}a}_0 + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\mathbf{B}(i)\mathbf{H}^T\mathbf{y}(i)]\right\} \\
\propto \;& \exp[-\frac{1}{2}(\mathbf{a}-\mathbf{a}_*)^T\Sigma_*^{-1}(\mathbf{a}-\mathbf{a}_*)] \sim N((\mathbf{a}_*, (\mathbf{\Sigma}_*)) \qquad \text{(A.2)}
\end{aligned}
$$

where we called :

$$
\begin{aligned}
\mathbf{\Sigma}_*^{-1} &= [\mathbf{\Sigma_0^{-1}} + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\mathbf{B(i)H^THB(i)}] \\
\mathbf{\Sigma}_*^T\mathbf{a}_* &= [\mathbf{\Sigma_0^{-1}a}_0 + \frac{1}{\sigma^2}\sum_{k=0}^{M-1}\mathbf{B}(i)\mathbf{H}^T\mathbf{y}(i)].
\end{aligned}
$$

Derivation of (4.12)

$$
\begin{aligned}
p(\sigma^2|\mathbf{a}, \mathbf{X}, \mathbf{Y}) &= p(\mathbf{a}, \sigma^2, \mathbf{X}|\mathbf{Y})/p(\mathbf{a}, \mathbf{X}|\mathbf{Y}) \\
&\propto p(\mathbf{a}, \sigma^2, \mathbf{X}|\mathbf{Y}) \qquad \text{(A.3)} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{PM/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{k=0}^{M-1}\|\mathbf{y}(i) - \mathbf{HAx}(i)\|^2\right) \\
&\quad \cdot \left(\frac{1}{\sigma^2}\right)^{(\nu_0/2)+1}\exp\left(-\frac{\nu_0\lambda_0}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sigma^2}\right)^{((\nu_0+PM)/2)+1}\exp\left(-\frac{\nu_0\lambda_0 + s^2}{2\sigma^2}\right) \\
&\sim \chi^{-2}\left(\nu_0 + PM, \frac{\nu_0\lambda_0 + s^2}{\nu_0 + PM}\right). \qquad \text{(A.4)}
\end{aligned}
$$

Derivation of (4.14)

$$
\begin{aligned}
& p(x_k(i)=1|\mathbf{a},\sigma^2,\mathbf{X_{ki}},\mathbf{Y}) \\
= \;& p(\mathbf{a},\sigma^2,\mathbf{X}|\mathbf{Y})/p(\mathbf{a},\sigma^2,\mathbf{X}_{ki}|\mathbf{Y}) \\
\propto \;& p(\mathbf{a},\sigma^2,\mathbf{X}|\mathbf{Y}) \qquad \text{(A.5)} \\
\propto \;& p_k(i)\exp\left(-\frac{1}{2\sigma^2}\sum_{l=0}^{M-1}\|\mathbf{y}(l)-\mathbf{HAx}(l))\|^2\right) \\
\propto \;& p_k(i)\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}(i)-\mathbf{HAx}(i))\|^2\right) \qquad \text{(A.6)} \\
\Rightarrow \;& \frac{P(x_k(i)=1|\mathbf{a},\sigma^2,\mathbf{X_{ki}},\mathbf{Y})}{P(x_k(i)=0|\mathbf{a},\sigma^2,\mathbf{X_{ki}},\mathbf{Y})} \\
= \;& \frac{p_k(i)}{1-p_k(i)}\exp\left\{\frac{1}{2\sigma^2}[\|\mathbf{y}(i)-\mathbf{HA}(\mathbf{x}_k^0(i)-\mathbf{1}_k)\|^2-\|\mathbf{y}(i)-\mathbf{HA}(\mathbf{x}_k^0(i)+\mathbf{1}_k)\|^2]\right\} \\
= \;& \frac{p_k(i)}{1-p_k(i)}\exp\left\{\frac{2}{\sigma^2}(\mathbf{HA1}_k)^T[\mathbf{y}(i)-\mathbf{HAx}_k^0(i)]\right\} \\
= \;& \frac{p_k(i)}{1-p_k(i)}\exp\left\{\frac{2A_k}{\sigma^2}\mathbf{h}_k^T[\mathbf{y}(i)-\mathbf{HAx}_k^0(i)]\right\} \qquad \text{(A.7)}
\end{aligned}
$$

with $\mathbf{1}_k$ being the K-dimensional vector with all-zero entries except for the *kth* entry, which is 1.

APPENDIX B: DERIVATION FOR THE IMPULSIVE CASE

Derivation of (4.30)

$$
\begin{aligned}
& p\left(\mathbf{a}|\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X},\mathbf{Y}\right) \\
& = p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y})/p(\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \\
& \propto p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \qquad \text{(B.1)} \\
& \propto \exp\left\{-\frac{1}{2}\sum_{i=0}^{M-1}\left[\mathbf{y}(i)-\mathbf{HB}(i)\mathbf{a}\right]^T\mathbf{\Lambda}(i)^{-1}\left[\mathbf{y}(i)-\mathbf{HB}(i)\mathbf{a}\right]\right\} \\
& \quad .\ \exp\left[-\frac{1}{2}(\mathbf{a}-\mathbf{a}_0)^T\Sigma_0^{-1}(\mathbf{a}-\mathbf{a}_0)\right] \\
& \propto \exp\left\{-\frac{1}{2}\mathbf{a}^T\left[\mathbf{\Sigma_0^{-1}}+\sum_{i=0}^{M-1}\mathbf{B(i)H^T\Lambda}(i)^{-1}\mathbf{HB(i)}\right]\mathbf{a}+\mathbf{a}^T\left[\mathbf{\Sigma_0^{-1}a}_0+\sum_{i=0}^{M-1}\mathbf{B}(i)\mathbf{H}^T\mathbf{\Lambda}(i)^{-1}\mathbf{y}(i)\right]\right\} \\
& \propto \exp[-\frac{1}{2}(\mathbf{a}-\mathbf{a}_*)^H\Sigma_*^{-1}(\mathbf{a}-\mathbf{a}_*)] \sim N\left((\mathbf{a}_*,\mathbf{\Sigma}_*)\right) \qquad \text{(B.2)}
\end{aligned}
$$

where we called :

$$
\begin{aligned}
\mathbf{\Sigma}_*^{-1} &= [\mathbf{\Sigma_0^{-1}}+\sum_{i=0}^{M-1}\mathbf{B(i)H^T\Lambda}(i)^{-1}\mathbf{HB(i)}] \\
\mathbf{\Sigma}_*^T\mathbf{a}_* &= [\mathbf{\Sigma_0^{-1}a}_0+\sum_{i=0}^{M-1}\mathbf{B}(i)\mathbf{H}^T\mathbf{\Lambda}(i)^{-1}\mathbf{y}(i)].
\end{aligned}
$$

Derivation of (4.33)

$$
\begin{aligned}
p(\sigma_l^2|\mathbf{a},\sigma_{\bar{l}}^2,\epsilon,\mathbf{I},\mathbf{X},\mathbf{Y}) &= p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y})/p(\mathbf{a},\sigma_{\bar{l}}^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \\
&\propto p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \qquad \text{(B.3)} \\
&\propto \left(\frac{1}{\sigma_l^2}\right)^{(1/2)\sum_{i=0}^{M-1}n_l(i)}\exp\left\{-\frac{1}{2\sigma_l^2}\sum_{i=0}^{M-1}\sum_{j=0}^{P-1}[y_j(i)-\xi_j^T\mathbf{Ax}(i)]^2 1_{I_j(i)=l}\right\} \\
&\quad .\ \left(\frac{1}{\sigma_l^2}\right)^{(\nu_l/2)+1}\exp\left(-\frac{\nu_l\lambda_l}{2\sigma_l^2}\right) \\
&= \left(\frac{1}{\sigma_l^2}\right)^{((\nu_l)/2)+(1/2)\sum_{i=0}^{M-1}n_l(i)+1}\exp\left(-\frac{\nu_l\lambda_l}{2\sigma_l^2}\right)
\end{aligned}
$$

$$\sim \quad \chi^{-2}\left([\nu_l + \sum_{i=0}^{M-1} n_l(i)], \frac{\nu_l\lambda_l + s_l^2}{\nu_l + \sum_{i=0}^{M-1} n_l(i)}\right) \tag{B.4}$$

where we called

$$s_l^2 = \sum_{i=0}^{M-1}\sum_{j=0}^{P-1}[y_j(i) - \xi_j^T\mathbf{A}\mathbf{x}(i)]^2 1_{I_j(i)=l}.$$

Derivation of (4.35)

$$\begin{aligned}
& p(x_k(i)=1|\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X_{ki}},\mathbf{Y}) \\
= \quad & p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y})/p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}_{ki}|\mathbf{Y}) \\
\propto \quad & p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \qquad (B.5) \\
\propto \quad & p_k(i)\exp\left\{-\frac{1}{2}\sum_{l=0}^{M-1}[\mathbf{y}(l)-\mathbf{H}\mathbf{A}\mathbf{x}(l)]^T\mathbf{\Lambda}(l)^{-1}[\mathbf{y}(l)-\mathbf{H}\mathbf{A}\mathbf{x}(l)]\right\} \\
\propto \quad & p_k(i)\exp\left\{-\frac{1}{2}[\mathbf{y}(i)-\mathbf{H}\mathbf{A}\mathbf{x}(i)]^T\mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i)-\mathbf{H}\mathbf{A}\mathbf{x}(i)]\right\} \\
\Rightarrow \quad & \frac{P(x_k(i)=1|\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{X_{ki}},\mathbf{Y})}{P(x_k(i)=0|\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{X_{ki}},\mathbf{Y})} \\
= \quad & \frac{p_k(i)}{1-p_k(i)}\exp\{\frac{1}{2}([\mathbf{y}(i)-\mathbf{H}\mathbf{A}(\mathbf{x}(i)-\mathbf{1}_k)]^T\mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i)-\mathbf{H}\mathbf{A}(\mathbf{x}(i)-\mathbf{1}_k)] \\
& -[\mathbf{y}(i)-\mathbf{H}\mathbf{A}(\mathbf{x}(i)+\mathbf{1}_k)]^T\mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i)-\mathbf{H}\mathbf{A}(\mathbf{x}(i)+\mathbf{1}_k)])\} \\
= \quad & \frac{p_k(i)}{1-p_k(i)}\exp\left\{2(\mathbf{H}\mathbf{A}\mathbf{1}_k)^T\mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i)-\mathbf{H}\mathbf{A}\mathbf{x}_k^0(i)]\right\} \\
= \quad & \frac{p_k(i)}{1-p_k(i)}\exp\left\{2A_k\mathbf{h}_k^T\mathbf{\Lambda}(i)^{-1}[\mathbf{y}(i)-\mathbf{H}\mathbf{A}\mathbf{x}_k^0(i)]\right\}. \qquad (B.6)
\end{aligned}$$

Derivation of (4.36)

$$\begin{aligned}
& p(I_j(i)=l|\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I}_{ji},\mathbf{X},\mathbf{Y}) \\
= \quad & p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y})/p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I}_{ji},\mathbf{X}|\mathbf{Y}) \\
\propto \quad & p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \propto P[I_j(i)=l|\epsilon] \\
& .\frac{1}{\sqrt{\sigma_l^2}}\exp\left\{-\frac{1}{2\sigma_l^2}[y_j(i)-\xi_j^T(A)\mathbf{x}(i)]^2\right\} \qquad (B.7) \\
\Rightarrow \quad & \frac{P(I_j(i)=1|\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I_{ji}},\mathbf{X},\mathbf{Y})}{P(I_j(i)=2|\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I_{ji}},\mathbf{X},\mathbf{Y})}
\end{aligned}$$

$$= \frac{1-\epsilon}{\epsilon}\sqrt{\frac{\sigma_2^2}{\sigma_1^2}}\exp\left\{\frac{1}{2}[y_j(i)-\xi_j^T\mathbf{A}\mathbf{x}(i)]^2(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2})\right\}. \quad \text{(B.8)}$$

Derivation of (4.37)

$$\begin{aligned}
& p(\epsilon|\mathbf{a},\sigma_1^2,\sigma_2^2,\mathbf{I},\mathbf{X},\mathbf{Y}) \\
= \;& p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y})/p(\mathbf{a},\sigma_1^2,\sigma_2^2,\mathbf{I},\mathbf{X}|\mathbf{Y}) \\
\propto \;& p(\mathbf{a},\sigma_1^2,\sigma_2^2,\epsilon,\mathbf{I},\mathbf{X}|\mathbf{Y}) \propto p(\epsilon)p((I)|\epsilon) \quad \text{(B.9)} \\
\propto \;& \epsilon^{a_0-1}(1-\epsilon)^{b_0-1}\epsilon^{\sum_{i=0}^{M-1}n_2(i)}(1-\epsilon)^{\sum_{i=0}^{M-1}n_1(i)} \\
\sim \;& \text{Beta}\left\{a_0+\sum_{i=0}^{M-1}n_2(i), b_0+\sum_{i=0}^{M-1}n_1(i)\right\}. \quad \text{(B.10)}
\end{aligned}$$

APPENDIX C: DERIVATION FOR THE CAUCHY DISTRIBUTION

Derivation of (4.47)

$$\begin{aligned} p(\mathbf{a}|t,s,\mathbf{X},\mathbf{Y}) &= p(\mathbf{a},t,s,\mathbf{X}|\mathbf{Y})/p(t,s,\mathbf{X}|\mathbf{Y}) \\ &\sim p(\mathbf{a},t,s,\mathbf{X}|\mathbf{Y}) \\ &= p(\mathbf{Y}|\mathbf{a},t,s,\mathbf{X}).p(\mathbf{a},t,s,\mathbf{X}) \\ &= p(\mathbf{Y}|\mathbf{a},t,s,\mathbf{X}).p(\mathbf{a}|t,s,\mathbf{X}).p(t,s,\mathbf{X}) \\ &\sim p(\mathbf{Y}|\mathbf{a},t,s,\mathbf{X}).p(\mathbf{a}|t,s,\mathbf{X}) \qquad \text{(C.1)} \\ &= \frac{s}{\pi\left(s^2+\left[(\sum_{i=0}^{M-1}y(i)-\mathbf{HB(i)a}-t\right]^2\right)} \\ &\quad .\ \exp -\frac{1}{2}(\mathbf{a}-\mathbf{a_0})\mathbf{\Sigma}_0^{-1}(\mathbf{a}-\mathbf{a_0})^T \qquad \text{(C.2)} \end{aligned}$$

Derivation of (4.48)

$$\begin{aligned} p(t|\mathbf{a},s,\mathbf{X},\mathbf{Y}) &= p(t,\mathbf{a},s,\mathbf{X}|\mathbf{Y})/p(\mathbf{a},s,\mathbf{X}|\mathbf{Y}) \\ &\sim p(t,\mathbf{a},s,\mathbf{X}|\mathbf{Y}) \\ &= p(\mathbf{Y}|t,\mathbf{a},s,\mathbf{X}).p(t,\mathbf{a},s,\mathbf{X}) \\ &= p(\mathbf{Y}|t,\mathbf{a},s,\mathbf{X}).p(t|\mathbf{a},s,\mathbf{X}).p(\mathbf{a},s,\mathbf{X}) \\ &\sim p(\mathbf{Y}|t,\mathbf{a},s,\mathbf{X}).p(t|\mathbf{a},s,\mathbf{X}) \qquad \text{(C.3)} \\ &= \frac{s}{\pi\left(s^2+\left[(\sum_{i=0}^{M-1}y(i)-\mathbf{HA}x(i))-t\right]^2\right)} \\ &\quad .\ \exp -\frac{1}{2}(\mathbf{t}-\mu_\mathbf{t})\mathbf{\Sigma}_t^{-1}(\mathbf{t}-\mu_\mathbf{t})^T \qquad \text{(C.4)} \end{aligned}$$

Derivation of (4.49)

$$\begin{aligned} p(s|\mathbf{a},t,\mathbf{X},\mathbf{Y}) &= p(s,\mathbf{a},t,\mathbf{X}|\mathbf{Y})/p(\mathbf{a},t,\mathbf{X}|\mathbf{Y}) \\ &\sim p(s,\mathbf{a},t,\mathbf{X}|\mathbf{Y}) \end{aligned}$$

$$
\begin{aligned}
&= p(\mathbf{Y}|s,\mathbf{a},t,\mathbf{X}).p(s,\mathbf{a},t,\mathbf{X}) \\
&= p(\mathbf{Y}|s,\mathbf{a},t,\mathbf{X}).p(s|\mathbf{a},t,\mathbf{X}).p(\mathbf{a},t,\mathbf{X}) \\
&\sim p(\mathbf{Y}|s,\mathbf{a},t,\mathbf{X}).p(s|\mathbf{a},t,\mathbf{X}) \qquad \text{(C.5)} \\
&= \frac{s}{\pi(s^2+[(\sum_{i=0}^{M-1} y(i)-\mathbf{HA}x(i))-t]^2)} \\
&\quad .(\frac{1}{s})^{1+(\nu_0/2)} \exp\left(-\frac{\nu_0\lambda_0}{2s}\right) \qquad \text{(C.6)}
\end{aligned}
$$

Derivation of (4.50)

$$
p(x_k(i)=1|s,\mathbf{a},t,\mathbf{X_{ki}},\mathbf{Y}) \sim p_k(i).\frac{s}{\pi(s^2+[(\sum_{i=0}^{M-1} r(i)-\mathbf{HA}x(i))-t]^2)} \qquad \text{(C.7)}
$$

$$
\begin{aligned}
\frac{p(x_k(i)=1|s,\mathbf{a},t,\mathbf{X_{ki}},\mathbf{Y})}{p(x_k(i)=0|s,\mathbf{a},t,\mathbf{X_{ki}},\mathbf{Y})} &= \frac{p_k(i)}{1-p_k(i)} \\
&\quad .\frac{\left(s^2+\left[(\sum_{i=0}^{M-1} y(i)-\mathbf{HA}(x_k^0(i)+\mathbf{1}_k))-t\right]^2\right)}{\left(s^2+\left[(\sum_{i=0}^{M-1} r(i)-\mathbf{HA}(x_k^0(i)-\mathbf{1}_k))-t\right]^2\right)} \qquad \text{(C.8)}
\end{aligned}
$$

REFERENCES

1. Win, M. Z. and R. A. Scholtz, "Impulse radio: how it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36 - 38, February 1998.

2. Win, M. Z. and R. A. Scholtz, "Ultra-Wide Bandwidth Time-Hopping Spread-Spectrum Impulse Radio for Wireless Multiple-Access Communications," *IEEE Transactions on Communications*, vol. 48, no. 4, pp. 679 - 690, April 2000.

3. Win, M. Z., R. A. Scholtz, and M. A. Barnes, "Ultra-wide bandwidth signal propagation for indoor wireless communications," *IEEE International Conference on Communications*, vol. 1, pp. 56 - 60, June 1997.

4. Win, M. Z. and R. A. Scholtz, "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Communications Letters*, vol. 2, pp. 51 - 53, February 1998.

5. Bennett, C. L. and G. F. Ross, "Time-domain electromagnetics and its applications," *Proceedings of IEEE*, vol. 66, pp. 299 - 318, March 1978.

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

7. Verdú, S., "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Transactions on Information Theory*, vol. IT-32, pp. 85 - 96, January 1986.

8. Poor, H. V., "On parameter estimation in DS/CDMA formats," *Advances in Communications and Signal Processing*, W. A. Porter and S. C. Kak, Eds. New York: Springer-Verlag, 1989.

9. Steinberg, Y. and H. V. Poor, "Sequential amplitude estimation in multiuser com-

munications," *IEEE Transactions on Information Theory*, vol. 40, pp. 11 - 20, January 1994.

10. Xie, Z., C. K. Rushforth, R. T. Short, and T. K. Moon, "Joint signal detection and parameter estimation in multiuser communications," *IEEE Transactions on Communications*, vol. 41, pp. 1208 - 1216, August 1993.

11. Fawer, U. and B. Aazhang, "A multiuser receiver for code-division multiple-access communications over multipath channels," *IEEE Transactions on Communications*, vol. COMM-43, no. 2/3/4, pp. 1556 - 1565, Feb./Mar./Apr. 1995.

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

13. Iltis, R. A. and L. Mailaender, "An adaptive multiuser detector with joint amplitude and delay estimation," *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 774 - 785, June 1994.

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

15. Moon, T. K., Z. Xie, C. K. Rushforth, and R. T. Short, "Parameter estimation in a multiuser communication system," *IEEE Transactions on Communications*, vol. 42, pp. 2553 - 2560, Aug. 1994.

16. Strom, E. and S. Miller, "Asynchronous DS-CDMA systems: Low complexity near-far resistant receiver and parameter estimation," Technical Report, Dept. Elect. Eng., Univ. Florida, Gainesville, January 1994.

17. Wang, X. and R. Chen, "Adaptive bayesian multiuser detection for synchronous CDMA with Gaussian and impulsive noise," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 2013 - 2028, July 2000.

18. Aazhang, B. and H. V. Poor, "Peformance of DS/SSMA communications in impulsive channels - Part I: Linear correlation receivers," *IEEE Transactions on Communications*, vol. COMM-35, pp. 1179 - 1187, November 1987.

19. Aazhang, B. and H. V. Poor, "Peformance of DS/SSMA communications in impulsive channels - Part II: Hard-limiting correlation receivers," *IEEE Transactions on Communications*, vol. COMM-36, pp. 88 - 96, January 1988.

20. Aazhang, B. and H. V. Poor, "An analysis of nonlinear direct-sequence correlators," *IEEE Transactions on Communications*, vol. COMM-37, pp. 723 - 731, July 1989.

21. Arnold, S. F., "Gibbs sampling," in Handbook of Statistics, C. R. Rao, Ed. New York: Elsevier, vol. 9, pp. 599 - 625, 1993.

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

23. Casella, G. and E. I. George, "Explaining the Gibbs sampler," American Statistics, vol. 46, pp. 167 - 174, 1992.

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

25. Wang, X. and R. Chen, "Blind turbo equalization in Gaussian and impulsive noise," *IEEE Transactions on Vehicular Technology*, vol.50, no. 4, pp. 1092 - 1105, July 2001.

26. Xu, Z., P. Liu and J. Tang, "Blind Multiuser Detection For Impulse Radio UWB Systems," *IEEE Topical Conference*, pp. 453 - 454, October 2003.

27. Berrou, C. and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo codes," *IEEE Transactions on Communications*, vol. 44, October 1996.

28. Douillard, C., M. Jézéquel, and C. Berrou, "Iterative correction of intersymbol

interference: Turbo equalization," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 507 - 511, September-October 1995.

29. Poor, H. V. and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Transactions on Information Theory*, vol. IT-43, pp. 858 - 871, May 1997.

30. Poor, H. V., "Turbo multiuser detection: A primer," *Journal of Communications and Networks*, vol. 3, no. 3, pp. 196 - 201, 2001.

31. Poor, H. V., "Dynamic programming in digital communications: Viterbi decoding to turbo multiuser detection," *Journal of Optimization Theory and Applications*, vol. 115, no. 3, pp. 629 - 657, 2002.

32. Yang, Z. G. and X. Wang, "Blind turbo multiuser detection for long-code multipath CDMA," *IEEE Transactions on Communications*, vol. 50, no. 1, pp. 112 - 125, 2002.

33. Poor, H. V., "Iterative multiuser detection," *IEEE Signal Processing Magazine*, vol. 21, issue 1, pp. 81 - 88, January 2004.

34. Moher, M., "An iterative multiuser decoder for near-capacity communications," *IEEE Transactions on Communications*, vol. COMM-46, pp. 870 - 880, July 1998.

35. Reed, M. C., C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer, "Iterative multiuser detection for CDMA with FEC: Near single user performance," *IEEE Transactions on Communications*, vol. 46, pp. 1693 - 1699, December 1998.

36. Wang, X. and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Transactions on Communications*, vol. 47, July 1999.

37. Giannakis, G. B. and L. Yang, "Multistage Block-Spreading for Impulse Radio Multiple Access Through ISI Channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1767 - 1777, December 2002.

38. Foerster, J. R., "Channel modeling sub-committee report (final), tech. rep. p802.15-02/368r5-sg3a," Technical Report, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), December 2002.

39. Lee, P. M., *Bayesian statistics: An introduction,* 2nd edition, London: Arnold, 1997.

40. Lehmann, E. L. and G. Casella, *Theory of Point Estimation,* 2nd edition, New York: Springer-Verlag, 1998.

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

42. Gelfand, E. and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398 - 409, June 1990.

43. Middleton, D., "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Transactions on Information Theory*, vol. 45, pp. 1129 - 1149, May 1999.

44. Casarin, R., "Bayesian inference for mixtures of stable distributions," Ph.D. thesis, Dept. of Economics, University Ca' Foscari, Venice, 2003.

45. Buckle, D. J., "Bayesian inference for stable distributions," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 605 - 613, June 1995.

46. Devroye, L., "Random variate generation for multivariate unimodal densities," *ACM Transactions on Modeling and Computer Simulation*, vol. 7, no. 4, pp. 447 - 477, October 1997.

47. Leydold, J., "A rejection technique for sampling from log-concave multivariate distributions," *ACM Transactions on Modeling and Computer Simulation*, vol. 8, no. 3, pp. 254 - 280, July 1998.

48. Zoloratev, V. M., "On representation of stable laws by integrals," *Selected Translations in Mathematical Statistics and Probability*, no. 6, pp. 84 - 86, 1966.

49. Samorodnitsky, G. and M.S. Taqqu, "Stable non-Gaussian random processes: stochastic models with infinite variance," Chambrdige University Press, 1994.

50. Adler, R. J., R.E. Feldman, M.S. Taqqu, "A practical guide to heavy tails: statistical techniques and applications," Birkauser Ed., Boston, 1998.

51. Bahl, L. R., J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimising symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284 - 287, March 1974.

52. Huber, P. J., *Robust Statistics*, New York: Wiley, 1981.

53. Huang, H. C. and S. Verdú, "Linear differentially coherent multiuser detection for multipath channels," *Wireless Personal Communications*, vol. 6, pp. 113 - 136, January 1998.