# STATISTICAL JOINT CHANNEL ESTIMATION AND DATA DETECTION FOR MULTIPLE ANTENNA WIRELESS COMMUNICATIONS

by

Xuehong Mao

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

The University of Utah

May 2013

**The University of Utah Graduate School**

# STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Xuehong Mao**

has been approved by the following supervisory committee members:

| | | |
|---|---|---|
| **Behrouz Farhang-Boroujeny** | , Chair | **12/14/2012** Date Approved |
| **Rong-rong Chen** | , Member | **12/14/2012** Date Approved |
| **Om P. Gandhi** | , Member | **12/14/2012** Date Approved |
| **Tolga Tasdizen** | , Member | **12/14/2012** Date Approved |
| **Osama S. Haddadin** | , Member | **12/19/2012** Date Approved |

and by **Gianluca Lazzi** , Chair of

the Department of **Electrical and Computer Engineering**

and by Donna M. White, Interim Dean of The Graduate School.

# ABSTRACT

Multiple-input and multiple-output (MIMO) technique has emerged as a key feature for future generations of wireless communication systems. It increases the channel capacity proportionate to the minimum number of transmit and receive antennas. This dissertation addresses the receiver design for high-rate MIMO communications in flat fading environments. The emphasis of the thesis is on the cases where channel state information (CSI) is not available and thus, clever channel estimation algorithms have to be developed to benefit from the maximum available channel capacity. The thesis makes four distinct novel contributions. First, we note that the conventional MCMC-MIMO detector presented in the prior work may deteriorate as SNR increases. We suggest and show through computer simulations that this problem to a great extent can be solved by initializing the MCMC detector with *regulated* states which are found through linear detectors. We also introduce the novel concept of *staged-MCMC* in a turbo receiver, where we start the detection process at a lower complexity and increase complexity only if the data could not be correctly detected in the present stage of data detection. Second, we note that in high-rate MIMO communications, joint data detection and channel estimation poses new challenges when a turbo loop is used to improve the quality of the estimated channel and the detected data. Erroneous detected data may propagate in the turbo loop and, thus, degrade the performance of the receiver significantly. This is referred to as error propagation. We propose a novel receiver that decorrelates channel estimation and the detected data to avoid the detrimental effect of error propagation. Third, the dissertation studies joint channel estimation and MIMO detection over a continuously time-varying channel and proposes a new dual-layer channel estimator to overcome the complexity of optimal channel estimators. The proposed dual-layer channel estimator reduces the complexity of the MIMO detector

with optimal channel estimator by an order of magnitude at a cost of a negligible performance degradation, on the order of 0.1 to 0.2 dB. The fourth contribution of this dissertation is to note that the Wiener filtering techniques that are discussed in this dissertation and elsewhere in the literature assume that channel (time-varying) statistics are available. We propose a new method that estimates such statistics using the coarse channel estimates obtained through pilot symbols. The dissertation also makes an additional contribution revealing differences between the MCMC-MIMO and LMMSE-MIMO detectors. We find that under the realistic condition where CSI has to be estimated, hence the available channel estimate will be noisy, the MCMC-MIMO detector outperforms the LMMSE-MIMO detector with a significant margin.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 Background and Motivation

Higher data rates and better link reliability are required by the ever-growing demand of wireless communications. The pioneering works by Winter, Telatar, Foschini, etc. [1–3] have inspired the wireless industry with the idea of using multiple antennas at both the transmit and receive side, i.e., multiple-input multiple-output (MIMO) communication. It has been shown that the capacity of a MIMO system increases linearly with the minimum of the number of transmit/receive antennas in a rich scattering environment [2]. This sets a strong theoretical foundation for MIMO technology. As a result, MIMO techniques have been widely incorporated in the current evolving wireless standards [4,5], such as 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE), Worldwide Interoperability for Microwave Access (WiMAX) for Wireless Metropolitan Area Networks (WMANs), IEEE 802.11n for Wireless Local Area Networks (WLANs), etc.

Although it has been claimed that the capacity gain (also referred to as *multiplexing gain*) of MIMO systems is up to the minimum number of the transmit/receive antennas, the bottleneck of such a system is the complexity of the receiver. In the perspective of computational cost, one of the most challenging tasks of the MIMO receiver is the MIMO detection, i.e., the tasks of resolving a large number of bits that are transmitted simultaneously per channel use. From this point of view, we propose to use the Markov chain Monte Carlo (MCMC) approach for MIMO detection, which is referred to as the MCMC-MIMO detector hereafter.

Another challenge in pursuing the multiplexing gain of MIMO systems is the reliable channel state information (CSI) at the receiver. Although CSI is required

by coherent MIMO receivers regardless of the detection algorithms, it can never be perfectly known at the receiver in reality. Therefore, an efficient channel estimation algorithm is desired in practical MIMO systems. The main task of this dissertation is to address channel estimation for large MIMO systems, i.e., the case where a large number of bits per channel use are being transmitted.

### 1.1.1 Theoretical Background of MIMO

Shannon capacity [6] is the most essential theoretical foundation of modern communications. It claims the maximum data rate that can be transmitted over the channel with an arbitrary small probability of error. It has been shown by Telatar in [2] that the capacity of MIMO channels increases linearly with the minimum of the number of transmit and receive antennas. Here, we attempt to give a brief discussion on the MIMO channel capacity. Throughout the discussion on capacity, let us consider the narrowband MIMO system shown in Fig.1.1. The MIMO system is expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1.1}$$

where $\mathbf{x} \in \mathcal{C}^{M\times 1}$ and $\mathbf{y} \in \mathcal{C}^{N\times 1}$ represent the input and output of the MIMO channel, respectively. $\mathbf{w} \in \mathcal{C}^{N\times 1}$ denotes the complex Gaussian noise with zero mean and unit variance (i.e., the covariance matrix of $\mathbf{w}$ is $\mathbf{C}_w = \mathrm{E}[\mathbf{w}\mathbf{w}^\dagger] = \mathbf{I}_N$). Assume there is a total power constraint $\rho$, i.e., $\sum_{m=1}^{M} \mathrm{E}[x_m x_m^*] = \rho$. Since the noise variance is 1, $\rho$ can also be interpreted as the total signal-to-noise ratio (SNR).

The MIMO channel in Fig.1.1 is represented by a $N \times M$ complex-valued matrix $\mathbf{H}$ with $h_{nm}$ representing the channel gain from transmit antenna $m$ to receive antenna $n$. This is consistent with the quasi-static channel model in our later discussions on MIMO channel modeling and estimation. To find out the MIMO channel capacity, different assumptions can be made on CSI. It is convenient to use CSIT and CSIR to represent the CSI known to the transmitter and receiver, respectively. When CSI is not available at either the transmit side or receive side, the zero-mean spatially white (ZMSW) model is the most common assumption. In the ZMSW model, entries of $\mathbf{H}$ are assumed to be independent and identically distributed (i.i.d.) zero mean, unit variance, and complex circularly symmetric Gaussian random variables. In general, different assumptions and knowledge on CSI lead to different MIMO channel

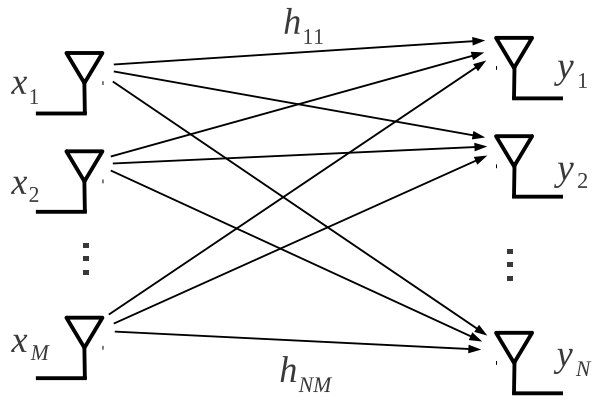


**Figure 1.1**. MIMO system model

capacities and space-time signaling approaches. For example, *waterfilling* is the optimum signaling strategy for CSIT and uniform power allocation is the optimum strategy for CSIR.

To show that the MIMO channel capacity increases linearly with $\min\{M, N\}$, we assume CSIT and obtain the singular value decomposition (SVD) of $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}, \tag{1.2}$$

where $\mathbf{U} \in \mathcal{C}^{N\times N}$ and $\mathbf{V} \in \mathcal{C}^{M\times M}$ are unitary matrices (i.e., $\mathbf{U}\mathbf{U}^{\dagger} = \mathbf{I}_N$, $\mathbf{V}^{\dagger}\mathbf{V} = \mathbf{I}_M$), and $\mathbf{\Sigma}$ is the diagonal matrix of the singular values of $\mathbf{H}$ (i.e., $\mathbf{\Sigma} = \text{diag}\{\sigma_1, \cdots \sigma_{R_H}\}$). Note that $R_H$ is the rank of $\mathbf{H}$, which also implies the number of non-zero singular values of $\mathbf{H}$. Since the rank of a matrix will never exceed the number of columns/rows, we have $R_H \leq \min\{M, N\}$. In a rich scattering environment, $\mathbf{H}$ will have full rank, i.e., $R_H = \min\{M, N\}$.

Substituting (1.2) in (1.1) and multiplying the result from the left by $\mathbf{U}^{\dagger}$, we obtain

$$\tilde{\mathbf{y}} = \mathbf{U}^{\dagger}\mathbf{y} = \mathbf{U}^{\dagger}(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}\mathbf{x} + \mathbf{w}) = \mathbf{\Sigma}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}, \tag{1.3}$$

where $\tilde{\mathbf{x}} = \mathbf{V}^{\dagger}\mathbf{x}$ and $\tilde{\mathbf{y}} = \mathbf{U}^{\dagger}\mathbf{y}$ are referred to as *transmit precoding* and *receiver shaping*. $\tilde{\mathbf{w}} = \mathbf{U}^{\dagger}\mathbf{w}$ and $\mathbf{w}$ have identical distribution, since multiplying by a unitary matrix does not change the distribution of a noise vector. Thus, the transmit precoding and receiver shaping convert the MIMO channel into $R_H$ parallel independent subchannels, as shown in Fig. 1.2.

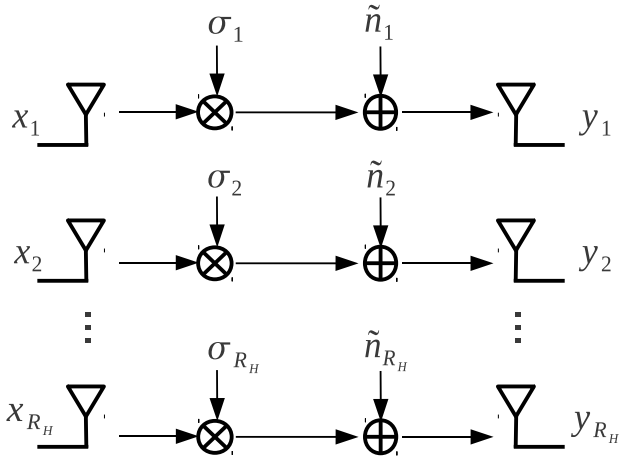


**Figure 1.2**. Parallel decomposition of MIMO channel

The parallel decomposition of the MIMO channel for a fixed **H** implies that MIMO capacity equals the sum of the capacities of each individual subchannel with the transmit power optimally allocated among these subchannels. Assume the bandwidth of the channel is $B$. Thus, we have

$$C = \max_{\rho_i : \sum_i \rho_i \leq \rho} \sum_i B \log_2 \left(1 + \rho_i \sigma_i^2\right) = \max_{\rho_i : \sum_i \rho_i \leq \rho} \sum_i B \log_2 \left(1 + \gamma_i\right). \tag{1.4}$$

Here, $\rho_i$ denotes the power allocated to the $i$-th subchannel. Solving the constrained optimization problem in (1.4) by the method of Lagrange multipliers, we find the waterfilling solution for the MIMO channel with CSIT as

$$\frac{\rho_i}{\rho} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma_i} & , \gamma_i \geq \gamma_0, \\ 0 & , \gamma_i < \gamma_0, \end{cases} \tag{1.5}$$

where $\gamma_i = \rho \sigma_i^2$ denotes the SNR associated with the $i$-th subchannel at full power and $\gamma_0$ is a cutoff value. The resulting capacity of MIMO with CSIT is

$$C = \sum_{i : \gamma_i \geq \gamma_0} B \log_2 \frac{\gamma_i}{\gamma_0}. \tag{1.6}$$

When the CSI is not available at the transmitter, i.e., in the case of CSIR, we derives the MIMO channel capacity in a different approach. This approach is based

on the definition of capacity, which is the maximized mutual information between the input and output of the MIMO channel. Mathematically, this is written as

$$C = \max_{p(\mathbf{x})} I(\mathbf{x};\mathbf{y}) = \max_{p(\mathbf{x})} \big(H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})\big). \tag{1.7}$$

Note that $H(\mathbf{y}|\mathbf{x})$ is the entropy of the noise $\mathbf{w}$ and thus is a fixed quantity. Hence, the capacity of the MIMO channel depends on the entropy of $\mathbf{y}$, which is determined by $\mathbf{C_y} = \mathrm{E}[\mathbf{y}\mathbf{y}^\dagger] = \mathbf{H}\mathbf{C_x}\mathbf{H}^\dagger + \mathbf{I}_N$. Similarly, $\mathbf{C_x}$ denotes the covariance matrix of the MIMO channel input. It has been shown that the entropy of $\mathbf{y}$ is maximized when $\mathbf{y}$ is a zero-mean circularly-symmetric complex Gaussian (ZMCSCG) random vector, which requires that the input $\mathbf{x}$ must be ZMCSCG as well. Thus, we have

$$C = \max_{p(\mathbf{x})} I(\mathbf{x};\mathbf{y}) = \max_{\mathbf{C_x}:\mathrm{Tr}\{\mathbf{C_x}\}=P} \log_2 \Big|\mathbf{I}_N + \mathbf{H}\mathbf{C_x}\mathbf{H}^\dagger\Big|. \tag{1.8}$$

Note that the capacity is achieved by maximizing the mutual information over $\mathbf{C_x}$ with the power constraint $\rho$. In the CSIR case, CSI is known to the receiver only. Thus, the waterfilling solution is not available because the transmitter cannot optimize the power allocation without knowing the singular values of $\mathbf{H}$. Intuitively, the best solution in this case is to transmit symbols from different antennas with the same power. The uniform power allocation leads to $\mathbf{C_x} = \frac{\rho}{M}\mathbf{I}_M$ and the mutual information

$$I(\mathbf{x};\mathbf{y}) = B\log_2 \Big|\mathbf{I}_N + \frac{\rho}{M}\mathbf{H}\mathbf{H}^\dagger\Big|. \tag{1.9}$$

By SVD, we find

$$I(\mathbf{x};\mathbf{y}) = \sum_{i=1}^{R_H} B\log_2\Big(1 + \frac{\gamma_i}{M}\Big), \tag{1.10}$$

where $\gamma_i = \rho\sigma_i^2$ as we defined previously.

Although MIMO promises the linear capacity growth with $\min\{M, N\}$, it has not specified how to extract such an attractive gain. In fact, the ultimate goal of this thesis is to develop a MIMO system that approaches this gain.

### 1.1.2 Spatial Multiplexing vs. Spatial Diversity

The core idea behind MIMO is the space-time signal processing where the data rate is maximized considering CSIT or CSIR approaches, and/or considering the

diversity of multilinks to improve the channel reliability. The first approach is called *spatial multiplexing*, while the second approach is referred to as*spatial diversity*.

Spatial multiplexing transmits independent symbols from each of the multiple transmit antennas and thus leads to an increase in the capacity. The idea behind spatial multiplexing is to exploit the spatial dimension, which is an additional source brought by multiple antennas. Assume the transmit antennas are located at different positions such that they will be assigned different spatial signatures by the MIMO channel. Thus, a receiver with multiple antennas can separate the different signals through their spatial signatures. For rich scattering channels with sufficiently largely separated antennas, spatial multiplexing provides a potential multiplexing gain up to the minimum of the number of transmit and receiver antennas. In [3], the Bell-Labs Layered Space Time (BLAST) high-speed wireless communication scheme was first reported to exploit the spatial multiplexing of MIMO technology. To extract the maximum multiplexing gain/degree of freedom, the transmitted data stream is split into $M$ substreams and launched by the $M$ transmit antennas simultaneously. The received signals are mixed spatially by the MIMO channel. At the receiver, the 'mixed' data have to be recovered by a suitable detection scheme.

Received signal power in a wireless channel fluctuates with time/frequency/space, which is called *fading*. Diversity is a technique which helps to stabilize a wireless link and combat fading. The basic idea is to create multiple copies of the transmitted signal for the receiver over the independent fading links. When the number of the independent links increases, the probability that all of them fade simultaneously decreases. In single antenna systems, diversity can be picked across time or frequency. In MIMO communications, diversity is available in an additional dimension - space. When the transmit/receive antennas are separated far enough to provide uncorrelated links, spatial diversity refers to the fact that the probability of losing the signal decreases exponentially with the number of uncorrelated links. To utilize the spatial diversity in MIMO systems, suitable code design is desired at the transmitter. Space-time code (STC), including space-time trellis code (STTC) [7] and space-time block code (STBC) [8], is proposed to transmit redundant copies of a data stream to the receiver. Although the data rate remains the same as in single-input single-output systems,

the transmission reliability is enhanced by spatial diversity.

In summary, MIMO systems can provide two types of gains: *spatial multiplexing gain* and *spatial diversity gain*. Spatial multiplexing is able to increase the data rate by extracting the multiplexing gain of the MIMO systems. On the other hand, spatial diversity utilizes the space-time processing in MIMO systems to combat fading to increase the reliability of wireless links. Also, it has been noted that there is a fundamental diversity-multiplexing trade-off [9]. Thus, a scheme maximizing one type of gain might not guarantee the other type of gain is maximized. In [9], Zheng and Tse verified that in the case of CSIR and a limited block length $T \geq M+N-1$ (i.e., quasi-static channel that does not change over $T$ symbol intervals), at asymptotically high SNR, the optimal diversity gain $d_{\text{opt}}$ achieved by a scheme with given multiplexing gain $r$ is

$$d_{\text{opt}}(r) = (M-r)(N-r), \ , 0 \leq r \leq \min\{M, N\}. \tag{1.11}$$

This implies that if we use all the transmit and receive antennas for diversity (i.e., $r = 0$), we may get the full diversity gain $d = M \cdot N$ and the error probability will be proportional to $\rho^{-M \cdot N}$.

Furthermore, Lozano and Jindal [10] argued that in the context of most modern wireless systems and for the operating points of interest, techniques utilizing the full degree of freedom for spatial multiplexing outperforms the spatial diversity techniques that explicitly sacrifice spatial multiplexing gain for spatial diversity gain. However, this conclusion might be violated in other cases if the channel model, performance metrics, and some key system features are chosen differently.

## 1.2 Overview of MIMO Systems

This dissertation addresses the problem of achieving high data rates in a computationally feasible manner. Thus, two technologies that are popular for the MIMO transmitter are adopted: BLAST architecture and bit-interleaved coded modulation (BICM).

In 1996, the first BLAST system was proposed by Foschini et al. [3] which has diagonal layering space-time coding with sequential nulling and interference cancelling decoding. This is also referred to as D-BLAST in later literatures. Although D-

BLAST is able to achieve the full spatial diversity gain, it suffers from the *boundary wastage* at the start and the end of each transmit packet, which becomes significant when the packet size is small. In 1997, vertical-BLAST (V-BLAST) [11] was stemmed from the work in [3]. V-BLAST overcomes the boundary wastage issue of D-BLAST by simply demultiplexing the transmitted data onto the different antennas without further preprocessing. However, the transmit diversity is vanished such that the diversity gain of V-BLAST will not exceed the number of receive antennas.

For fading channels, channel coding is desired to improve the reliability of wireless communication systems. BICM is a well-known coded modulation strategy that combines the channel coding and symbol mapping through a bit-wise interleaver. With a soft-input channel decoder, BICM yields excellent performance over Rayleigh flat-fading channels. The performance can be further improved if *turbo-principle* is applied such that the data detection (demapping) and channel decoding are operated in an iterative manner. Here, turbo principle refers to a general approach for combining and serially performing two or more tasks of the receiver in an iterative manner. The idea stemmed from *turbo codes* [12] founded by Berrou et al. and has been extended to many fields other than channel coding theory. For MIMO systems, BICM necessitates a turbo-loop between the soft-input soft-output (SISO) MIMO-detector and channel decoder. Thus the overall system performance can be improved significantly.

To give an overview of MIMO systems and make this dissertation self-contained, the most related techniques are introduced in this section. Let us start with BLAST architectures and their detection algorithms.

### 1.2.1 The V-BLAST Architecture

The V-BLAST [11] is a layered space-time architecture originally proposed and implemented by Bell-Labs to achieve high data rates promised by MIMO technology. Fig. 1.3 illustrates a high-level block diagram of a V-BLAST system.

In this scheme, a single data stream is demultiplexed into $M$ substreams, denoted by $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$, and each substream is then fed to its respective transmitter. The $M$ independent substreams are transmitted simultaneously by all transmitter antennas (i.e., they share both frequency and time). Each received signal is a superposition

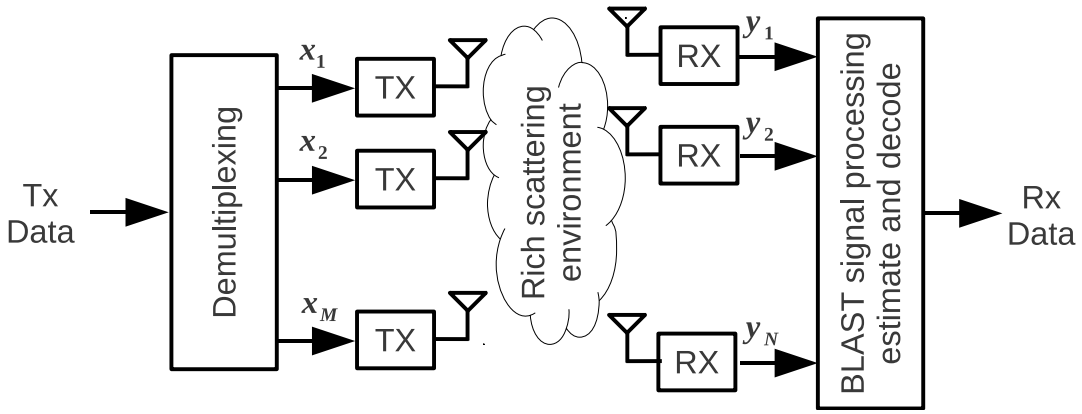


**Figure 1.3**. V-BLAST high-level system diagram

of the signals transmitted by all $M$ antennas. By appropriate signal processing, which is called MIMO detection, in general, individual symbols are recovered. The discrete-time baseband model of the V-BLAST system can be expressed as

$$\mathbf{y} = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1.12}$$

where $\mathbf{y} \in \mathcal{C}^{N\times 1}$ and $\mathbf{x} \in \mathcal{C}^{M\times 1}$ denote the received and transmitted symbols, respectively. Each symbol in $\mathbf{x}$ is obtained from a finite constellation $\mathcal{A}$. Since we assume that the CSI is not available at the transmitter, uniform power allocation is adopted with a total power power constraint: $\mathrm{E}[x_m x_m^*] = 1, \ , 1 \leq m \leq M$. Note that the representation of (1.12) is the same as the that of (1.1) except for the scaling factor $\sqrt{\frac{\rho}{M}}$ due to the normalized power for the symbol constellation. $\mathbf{w} \in \mathcal{C}^{N\times 1}$ is the noise vector with components drawn from an i.i.d $w_i \sim \mathcal{N}(0,1)$. $\mathbf{H} \in \mathcal{C}^{N\times M}$ denotes the channel matrix, where $h_{i,j}$ is the complex channel gain between the $j$-th transmit antenna and the $i$-th receive antenna and $\rho$ is interpreted as the total signal-to-noise ratio (SNR). Note that (1.12) is a narrowband baseband model. In the case of a wideband system, OFDM can be used to obtain a set of parallel narrowband sub-MIMO systems. Thus, presentation in (1.12) can be viewed as a snapshot of the OFDM-MIMO system at a particular frequency (subcarrier) and at a specific instant of time.

In the following, assuming the channel matrix $\mathbf{H}$ is perfectly known at the receiver, we discuss the various detection techniques that have been proposed for V-BLAST.

#### 1.2.1.1 Maximum Likelihood (ML) Detector

The ML detector is the optimal receiver in terms of minimizing the bit error rate (BER). For the system defined in (1.12), the ML detector is given by

$$\hat{\mathbf{x}}_{\text{ml}} = \underset{\mathbf{x}\in\mathcal{A}^M}{\arg\min} \left\| \mathbf{y} - \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{x} \right\|^2. \tag{1.13}$$

The minimizing problem in (1.13) requires an exhaustive search over all possible vectors of transmitted symbols. The computational complexity of such an exhaustive search is $\mathcal{O}({M_c}^M)$ with $M_c = |\mathcal{A}|$ denoting the constellation size. As a result, the ML detector results in an exponentially growing complexity with the number of transmit antennas. This undesired feature of the ML detector is dealt with using various low-complexity near-optimal approaches. Sphere decoding (SD) [13] is such an algorithm with $\mathcal{O}({M_c}^3)$ complexity founded by binary tree-search theory. On the other hand, the Monte Carlo Markov chain (MCMC) detector, ignited in stochastic approximation, is of interest to me. More detailed discussion on the MCMC detector will be provided in Chapter 3.

#### 1.2.1.2 Sphere Detector

Due to the exponentially growing complexity of the ML detector, suboptimal detectors with lower complexity are highly demanded. The sphere detector (SD), also referred to as sphere decoder, is an suboptimal solution to ML detector which avoids the exhausting search by examining the points that lie inside a hypersphere.

$$(\mathbf{x} - \hat{\mathbf{x}})^\dagger \mathbf{H}^\dagger \mathbf{H} (\mathbf{x} - \hat{\mathbf{x}}) \leq r^2, \tag{1.14}$$

where $\hat{\mathbf{x}} = \sqrt{\frac{M}{\rho}}(\mathbf{H}^\dagger\mathbf{H})^{-1}\mathbf{H}^\dagger\mathbf{y}$ is the constrained ML estimate of $\mathbf{x}$ and $r$ is the pre-defined radius of the sphere. On one hand, $r$ has to be large enough to contain the true ML estimate which is found by

$$\hat{\mathbf{x}}_{\text{ml}} = \underset{\mathbf{x}\in\Lambda}{\arg\min} (\mathbf{x} - \hat{\mathbf{x}})^\dagger \mathbf{H}^\dagger \mathbf{H} (\mathbf{x} - \hat{\mathbf{x}}), \tag{1.15}$$

where $\Lambda$ is the lattice defined by having each $M$-dimensional element $\mathbf{x}$ taken from a constellation of $2^{M_c}$ consecutive integers. On the other hand, $r$ should be as small as possible such that the hypersphere only contains a few candidates.

#### 1.2.1.3 Linear Detectors

Linear detectors are a class of suboptimal MIMO detectors with low (linear) complexity. In general, they are operated by applying a certain type of filter, denoted by $\mathbf{G} \in \mathcal{C}^{M \times N}$, to the received signal $\mathbf{y}$.

• **Zero-forcing (ZF) detector** To recover the desired signal from each transmitted antenna, the ZF detector treats the signals from other antennas as interference and suppresses them by a filter $\mathbf{G}_{\text{zf}}$

$$\mathbf{G}_{\text{zf}} = \sqrt{\frac{\rho}{M}}(\mathbf{H}^{\dagger}\mathbf{H})^{-1}\mathbf{H}^{\dagger}, \tag{1.16}$$

which is a pseudo-inverse of the channel matrix with appropriate scaling. The output of the ZF sector is

$$\mathbf{z}_{\text{zf}} = \mathbf{x} + \mathbf{G}_{\text{zf}}\mathbf{w}. \tag{1.17}$$

Note that, here, the desired signal $\mathbf{x}$ is resolved, whereas the noise term $\mathbf{w}$ is multiplied by $\mathbf{G}_{\text{zf}}$, which may lead to noise enhancement. Therefore, the poor performance at the low SNR region is a known disadvantage of the ZF detector.

• **Minimum mean square error (MMSE) detector** The noise enhancement problem of the ZF detector can be solved by the MMSE detector, where

$$\mathbf{G}_{\text{mmse}} = \sqrt{\frac{\rho}{M}}(\mathbf{H}^{\dagger}\mathbf{H} + \frac{M}{\rho}\mathbf{I}_M)^{-1}\mathbf{H}^{\dagger}. \tag{1.18}$$

The MMSE detector minimizes the mean square error (MSE) $\text{E}[\|\mathbf{G}_{\text{mmse}}\mathbf{y} - \mathbf{x}\|^2]$. Although the MMSE detector outperforms the ZF detector at low SNR, their performance converges to the same point, which exploits a diversity of order $M-N+1$ [14]. None of them could achieve the full diversity $\min\{M, N\}$.

#### 1.2.1.4 Nulling and Canceling (NC) Detector

In contrast to the linear detectors, the NC detector uses a sequential decision-feedback approach to detect the symbol layers one after another. Although it promises better diversity order than the pure linear detector, the NC detector suffers from

**Algorithm 1.1:** MMSE-NC with ordering

**for** $i = 1$ *to* $M$ **do**
- $\mathbf{G}_{mmse} = \sqrt{\frac{\rho}{M}}(\mathbf{H}^{\dagger}\mathbf{H} + \frac{M}{\rho}\mathbf{I}_M)^{-1}\mathbf{H}^{\dagger}$ (MMSE filter);
- $k_i = \arg\min_{j \neq i}\{g_{j,j}\}$ (ordering);
- $z = \mathbf{G}_{mmse}(k_i, :)\mathbf{H}^{\dagger}\mathbf{y}$ (nulling);
- $\mathbf{y} = \mathbf{y} - \sqrt{\frac{\rho}{M}}\mathbf{H}(:, k_i)\mathcal{Q}(z)$ (quantization and canceling) ;
- removing the $k_i$-th column of $\mathbf{H}$;

**end**

the error propagation, i.e., an incorrect decision in the detection of a symbol adds interference to the next symbol to be detected. The ordered NC detector [11] is a improved version of NC detector. It detects the symbol with the highest signal-to-interference-plus-noise ratio (SINR) first, and proceeds with the detection of the next symbol in the same manner. This algorithms is summarized in Algorithm 1.1.

### 1.2.2 Soft-input Soft-output (SISO) Detectors

To combat fading channels, wireless communication systems usually use channel coding to improve the reliability of wireless communication systems. For such a *coded system*, the optimal MIMO detector has to make decisions jointly on all the coded bits in one code block using the correlations introduced by the channel encoder. Similarly, the channel decoder should also take the likelihood information on all coded bits into account. Generally, the complexity of such a joint detection/decoding is prohibitive even with reasonable coding length. BICM and turbo principle are effective means to solve this problem in an iterative manner.

A block diagram of the transceiver structure of the proposed system is shown in Fig.1.4. An information bit sequence $\mathbf{b}$ is encoded by a channel encoder (e.g., convolutional code) of rate $R$. The coded bit sequence $\mathbf{d}$ is interleaved according to a permutation function $\Pi(\cdot)$. Then every $M_c = \log_2 |\mathcal{A}|$ consecutive interleaved bits are partitioned into a group and mapped to a complex-valued symbol on the constellation $\mathcal{A}$. After inserting pilot symbols, the resulting symbol sequence is formed to a signal matrix $\mathbf{X} \in \mathcal{C}^{M \times T}$ and sent through the MIMO channel. Here, the first $T_p$ columns of $\mathbf{X}$ consist of pilot symbols, and the rest of $T - T_p$ columns consist of data symbols. Although the transmitted signal and received signal are represented by matrices in

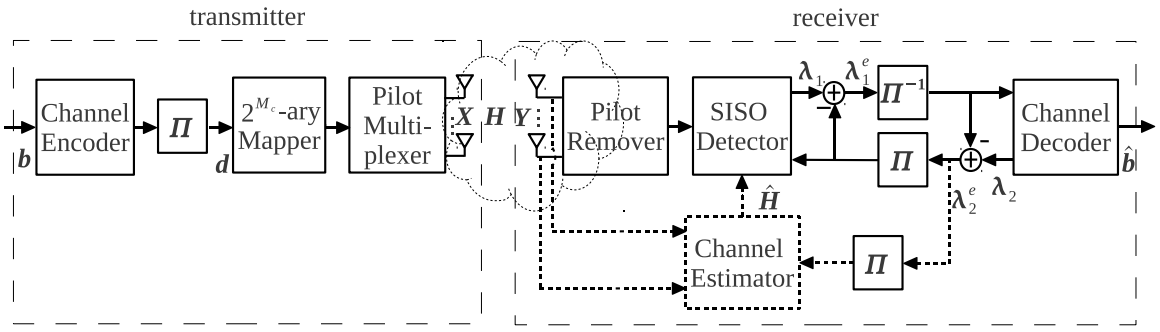


**Figure 1.4**. BICM-MIMO system diagram

Fig. 1.4, we may treat them as a result of applying (1.12) $T$ times.

As illustrated in Fig. 1.4, the receiver is operated in an iterative manner. It consists of three operating modules, SISO detector, channel decoder, and channel estimator (optional). The soft information, in terms of the *a posteriori probabilities* (APPs), is interchanged between the three modules. We let $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ denote the output APPs of the code bits produced by the SISO detector and channel decoder, respectively. The corresponding symbol-wise APPs are represented by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ (not shown in the figure). Note that we use subscript "1" and "2" to distinguish the APP from the MIMO detector and the channel decoder. To prevent the error propagation, extrinsic information (after subtracting the *a priori probabilities* in log domain) $\boldsymbol{\lambda}_1^e$ and $\boldsymbol{\lambda}_2^e$ are exchanged between decoder/detector. At the first iteration, channel estimate is initialized from the pilot symbols and we term this as *training mode* in the later parts of this thesis. In the successive iteration, channel is refined by using $\boldsymbol{\lambda}_2$ and $\mathbf{Y}$. Subsequently, $\hat{\mathbf{H}}$ and $\boldsymbol{\lambda}_2^e$ are fed to the SISO detector for data detection. The SISO detector generates updated symbol probabilities $\boldsymbol{\lambda}_1$, and the extrinsic information $\boldsymbol{\lambda}_1^e$ is passed back to the channel decoder for data decoding. In this way, joint SISO detection, channel estimation, and data decoding is performed iteratively. After a predetermined number of iterations, decisions are made at the receiver output to obtain the estimated information bit sequence $\hat{\mathbf{b}}$.

The MIMO detector is a key component at the receiver in a BICM-MIMO system.

The task of the SISO detector is to provide the soft information of the transmitted symbols $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_M]^{\mathrm{T}}$ from the revived signal $\mathbf{y}$ and the channel matrix $\mathbf{H}$ or its estimate $\hat{\mathbf{H}}$. For simplicity, we assume channel gain $\mathbf{H}$ is given in this section.

#### 1.2.2.1 Symbol-wise SISO (*s*-SISO) detector

Instead of obtaining explicit symbol values, the $s$-SISO detector computes the APPs of each transmitted symbol $x_m$, $1 \leq m \leq M$ according to the maximum $a$ *posteriori* (MAP) criterion

$$\mu_1(x_{m,j}) \triangleq P(x_m = a_j|_{\mathbf{y},\mathbf{H}}) = \sum_{\mathbf{x}_{-m} \in \mathcal{A}^{M-1}} P(x_m = a_j, \mathbf{x}_{-m}|_{\mathbf{y},\mathbf{H}}), \tag{1.19}$$

where $a_j$ $(1 \leq j \leq 2^{M_c})$ is the $j$-th constellation point. $\mathbf{x}_{-m}$ denotes the $(M-1)$-dimensional vector of transmitted symbols excluding the one from the $m$-th antenna. Applying Bayes' rule to (1.19), we find

$$\begin{aligned} \mu_1(x_{m,j}) &= \sum_{\mathbf{x}_{-m} \in \mathcal{A}^{M-1}} p(\mathbf{y}|_{x_m = a_j, \mathbf{x}_{-m}, \mathbf{H}}) P(x_m = a_j, \mathbf{x}_{-m}) p(\mathbf{H}) \\ &= p(\mathbf{H}) \cdot P(x_m = a_j) \cdot \sum_{\mathbf{x}_{-m} \in \mathcal{A}^{M-1}} p(\mathbf{y}|_{x_m = a_j, \mathbf{x}_{-m}}) P(\mathbf{x}_{-m}). \end{aligned} \tag{1.20}$$

In (1.20), $p(\mathbf{H})$ is the probability density function (PDF) of a given channel realization $\mathbf{H}$ such that it can be modeled as a common factor for all $j \in [0, 2^{M_c}]$. $P(x_m = a_j) = \mu_2^e(x_{m,j})$ and $P(\mathbf{x}_{-m})$ are symbol probabilities computed from extrinsic APP obtained from the channel decoder. Note that we use $p(\cdot)$ and $P(\cdot)$ to distinguish the PDF for continuous variables and the probability mass function (PMF) for discrete variables, respectively. Thus, we have

$$\mu_1(x_{m,j}) \propto \mu_2^e(x_{m,j}) \cdot \sum_{\mathbf{x}_{-m} \in \mathcal{A}^{M-1}} p(\mathbf{y}|_{x_m = a_j, \mathbf{x}_{-m}}) P(\mathbf{x}_{-m}). \tag{1.21}$$

For the iterative receiver, it is convenient to compute the APP in log-domain. By taking the logarithm of (1.21) on both sides, it is obvious that the extrinsic APP required by the channel decoder is the logarithm of the second multiplier on the right-hand side of (1.21).

#### 1.2.2.2 Bit-wise SISO (*b*-SISO) detector

The SISO detector can also be operated in bit-wise, which aims at computing the log-likelihood ratio (LLRs) of the transmitted bits. Let $x_m = \mathcal{M}(\mathbf{d}_m)$ denote

the symbol mapping operator, where $\mathbf{d}_m = [d_{m,1} \quad d_{m,2} \quad \cdots \quad d_{m,M_c}]$ denotes the bits consisting of $x_m$. We assume that the coded bits $d_{m,k}$ are equally likely and statistically independent, which is a good approximation for a sufficiently long coding and interleaving length. The APP of $d_{m,k}$, $, 1 \leq k \leq M_c$ is defined as

$$\lambda_1(d_{m,k}) \triangleq \frac{p(d_{m,k}=1|_{\mathbf{y},\mathbf{H}})}{p(d_{m,k}=0|_{\mathbf{y},\mathbf{H}})} = \frac{\sum_{\mathbf{d}\in\mathcal{D}^1_{m,k}} P(d_{m,k}=1,\mathbf{d}_{-m,k}|_{\mathbf{y},\mathbf{H}})}{\sum_{\mathbf{d}\in\mathcal{D}^0_{m,k}} P(d_{m,k}=0,\mathbf{d}_{-m,k}|_{\mathbf{y},\mathbf{H}})}, \tag{1.22}$$

where $\mathbf{d}_{-m,k}$ denotes the coded bit vector excluding $d_{m,k}$ and $\mathcal{D}^b_{m,k}$ denotes the set of all possible transmitted bit vector $\mathbf{d} = [d_{1,1} \quad \cdots \quad d_{1,M_c} \quad \cdots \quad d_{M,1} \quad \cdots \quad d_{M,M_c}]$ whose $m$-th element at bit position $k$ equals $b \in \{0,1\}$, i.e.,$\mathcal{D}^b_{m,k} \triangleq \{\mathbf{d}|d_{m,k}=b\}$. Applying Bayes' rule to (1.22), we find

$$\begin{aligned}\lambda_1(d_{m,k}) &= \frac{P(d_{m,k}=1)\cdot\sum_{\mathbf{d}\in\mathcal{D}^1_{m,j}} p(\mathbf{y}|_{\mathbf{x}=\mathcal{M}(\mathbf{d}),\mathbf{H}})P(\mathbf{d}_{-m,k})}{P(d_{m,k}=0)\cdot\sum_{\mathbf{d}\in\mathcal{D}^0_{m,j}} p(\mathbf{y}|_{\mathbf{x}=\mathcal{M}(\mathbf{d}),\mathbf{H}})P(\mathbf{d}_{-m,k})} \\ &= \lambda^e_2(d_{m,k})\cdot\frac{\sum_{\mathbf{d}\in\mathcal{D}^1_{m,j}} p(\mathbf{y}|_{\mathbf{x}=\mathcal{M}(\mathbf{d}),\mathbf{H}})P(\mathbf{d}_{-m,k})}{\sum_{\mathbf{d}\in\mathcal{D}^0_{m,j}} p(\mathbf{y}|_{\mathbf{x}=\mathcal{M}(\mathbf{d}),\mathbf{H}})P(\mathbf{d}_{-m,k})}.\end{aligned} \tag{1.23}$$

Based on (1.21) and (1.23) , the complexity of the optimal maximum *a posteriori* MAP detector increases exponentially in $M$ because there are $2^{(M-1)M_c}$ terms in the summation.

### 1.2.3 Soft-MMSE MIMO Detector

To reduce the complexity of the MAP MIMO detector and harvest the gain brought by the turbo principle, the soft-MMSE MIMO detector is proposed in the literature. It originates from [15], which proposed an equalizer using *a priori* information for single antenna systems. A similar idea has been presented in [16]. The details of the soft-MMSE MIMO detector is summarized as follows.

Consider a V-BLAST MIMO system shown in (1.12). A linear estimator of the transmitted symbol $x_m (1 \leq m \leq M)$ using the observation $\mathbf{y}$ is given by

$$\hat{x}_m = \mathbf{a}^\dagger_m \mathbf{y} + b_m, \tag{1.24}$$

where $\mathbf{a}_m \triangleq [a_1 \quad \cdots \quad a_N]^{\mathrm{T}} \in \mathcal{C}^{N\times 1}$ and $b_m \in \mathcal{C}$ are the coefficients of the estimator. The optimal coefficients are given by

$$\mathbf{a}_m = \mathrm{E}[\mathbf{y}\cdot\mathbf{y}^\dagger]^{-1}\mathrm{E}[\mathbf{y}\cdot x_m], \tag{1.25a}$$

and

$$b_m = \mathrm{E}[x_m] - \mathbf{a}_m^\dagger \mathrm{E}[\mathbf{y}], \tag{1.25b}$$

which minimizes the MSE [17]. Given the *a priori* information $\boldsymbol{\mu}$, the mean and variance of each transmitted symbol $x_m$, denoted by $\bar{x}_m$ and $v_m$, respectively, are obtained as

$$\bar{x}_m = \sum_{a \in \mathcal{A}} a \cdot P(x_m = a|\mu_m), \quad v_m = \sum_{a \in \mathcal{A}} |a|^2 \cdot P(x_m = a|\mu_m) - |\bar{x}_m|^2, \tag{1.26}$$

where $\mathcal{A}$ denotes the set of data symbol constellation points. Using $\bar{x}_m$ and $v_m$ to define

$$\mathrm{E}[\mathbf{y}] = \sqrt{\frac{\rho}{M}} \mathbf{H} \bar{\mathbf{x}}, \tag{1.27a}$$

$$\mathrm{E}[\mathbf{y} \cdot x_m] = \sqrt{\frac{\rho}{M}} v_m \mathbf{h}_m, \tag{1.27b}$$

$$\mathrm{E}[\mathbf{y} \cdot \mathbf{y}^\dagger] = \mathbf{I}_N + \frac{\rho}{M} \mathbf{H} \mathbf{V} \mathbf{H}^\dagger \triangleq \boldsymbol{\Sigma}_m, \tag{1.27c}$$

where $\bar{\mathbf{x}} = [\bar{x}_1 \quad \cdots \quad \bar{x}_M]^{\mathrm{T}}$, $\mathbf{V} = \mathrm{diag}[v_1 \quad \cdots \quad v_M]^{\mathrm{T}}$, and $\mathbf{h}_m$ is the $m$-th column of the channel matrix $\mathbf{H}$. Thus, the MMSE estimate $\hat{x}_m$ is given by

$$\hat{x}_m = \bar{x}_m + \sqrt{\frac{\rho}{M}} v_m \mathbf{h}_m^\dagger \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \sqrt{\frac{\rho}{M}} \mathbf{H} \bar{\mathbf{x}}). \tag{1.28}$$

Note that $\hat{x}_m$ depends on $\mu_m$ via $\bar{x}_m$ and $v_m$. To make $\hat{x}_m$ independent from $\mu_m$, we set $\bar{x}_m = 0$ and $v_m = 1$ in (1.28) and obtain the estimates $\hat{x}_m$ $(1 \le m \le M)$ as

$$\hat{x}_m = \sqrt{\frac{\rho}{M}} \mathbf{h}_m^\dagger (\boldsymbol{\Sigma}|_{v_m=1})^{-1} (\mathbf{y} - \sqrt{\frac{\rho}{M}} \mathbf{H} \bar{\mathbf{x}}|_{\bar{x}_m=0}). \tag{1.29}$$

By the Gaussian assumption [15], $p(\hat{x}_m|_{x_m=a_j})$, $j = 1, \cdots, 2^{M_c}$ are Gaussian random variables with the mean $\nu_{m,j} = \mathrm{E}[\hat{x}_m|_{x_m=a_j}]$ and the variance $\sigma_{m,j} = \mathrm{E}[\hat{x}_m \hat{x}_m^*|_{x_m=a_j}]$. Using $\bar{x}_m$ and $v_m$, the mean and variance of $p(\hat{x}_m|_{x_m=a_j})$ are computed by

$$\nu_{m,j} = a_j K_m \sqrt{\frac{\rho}{M}} \mathbf{f}_m^\dagger \mathbf{h}_m, \tag{1.30a}$$

and

$$\sigma_{m,j}^2 = K_m^2 (\sqrt{\frac{\rho}{M}} \mathbf{f}_m^\dagger \mathbf{h}_m - v_m \frac{\rho}{M} \mathbf{f}_m^\dagger \mathbf{h}_m \mathbf{h}_m^\dagger \mathbf{f}_m), \tag{1.30b}$$

where $\mathbf{f}_m = \sqrt{\frac{\rho}{M}} \boldsymbol{\Sigma}_m^{-1} \mathbf{h}_m$ and $K_m = \frac{1}{1 + (1 - v_m)\sqrt{\frac{\rho}{M}} \mathbf{f}_m^\dagger \mathbf{h}_m}$.

Finally, the soft-MMSE MIMO detector computes the LLR values of $x_m$ ($1 \leq m \leq M$) by

$$\mu_1(x_{m,j}) = \frac{1}{\pi\sigma_{m,j}^2} \exp\Big\{ - \frac{|\bar{x}_m - \nu_{m,j}|^2}{\sigma_{m,j}^2} \Big\}. \tag{1.31}$$

Compared with the symbol-wise MAP SISO MIMO detector, the soft-MMSE MIMO detector is a suboptimal detector whose complexity grows linearly with the number of transmit antennas. In the meantime, matrix inversion is required while obtaining the MMSE coefficients. The performance of the soft-MMSE MIMO detector is compared with the MCMC MIMO detector in later chapters under various situations.

## 1.3 Dissertation Outline and Contributions

Chapter 2 provides an overview of channel modeling and estimation in wireless communications. Since dealing with the radio propagation is one of the challenges in practice, this chapter presents typical mathematical/statistical channel models that are widely adopted in the literature. The models developed in this chapter will be used in the later chapters of this thesis. Given a model of the radio propagation, channel estimation algorithms are developed to acquire/track the behavior of the channel. An overview of the conventional channel estimation approaches is presented in Chapter 2.

Chapter 3 introduces the basic MCMC-MIMO detector. At the beginning of this chapter, general Monte Carlo integration and importance sampling are presented. The basic properties of the Markov chain are introduced as the background of the MCMC method as well. Two well-known MCMC algorithms, Metropolis-Hasting and Gibbs sampling, are then presented. The detail of the MCMC-MIMO detector and a dilemma of the MCMC detector are discussed in the later parts of the chapter.

In Chapter 4, we assume perfect CSI is known at the receiver. Two solutions are proposed to alleviate performance degradation of the MCMC-MIMO detector in the high SNR regimes. One is *regulated-MCMC*, and the other is *staged-MCMC*. Instead of random initialization, regulated-MCMC initializes MCMC with some deterministic states found through a simple linear detector. Staged-MCMC is a novel concept in a turbo receiver. It starts the detection process at a lower complexity and increases

complexity only if the data are not correctly detected in the present stage of data detection. The effectiveness of the proposed methods is shown through empirical (simulation) results.

Chapter 5 emphasizes receiver design for the high-rate MIMO system over the block fading channel. A novel decorrelation receiver is proposed. This method breaks the correlation between soft-decision-directed channel estimation (SCE) and MIMO detection, to achieve an excellent performance close to that of a genie-aided receiver. The new design is applicable to systems with a large number of transmit antennas, arbitrary modulation size, unknown fading correlations, and arbitrary detection methods. It also provides us a more realistic performance benchmark when applying the decorrelation design to genie-aided channel estimation. By simulations, it is demonstrated that with the decorrelation design, MCMC-MIMO detector is superior to the state-of-the-art LMMSE-MIMO detector.

Chapter 6 addresses receiver design for the high-rate MIMO system over time-varying fading channels. To deal with the high complexity issue of the optimal channel estimator, a novel dual-layer channel estimator is developed. In the first layer, a set of coarse channel estimates is obtained through a low complexity process. The second layer takes these coarse estimates and passes them through a smoothing (Wiener) filter that accounts for the channel variation with time. The proposed dual-layer channel estimator reduces the complexity of the MIMO detector by an order of magnitude at a cost of a negligible degradation. Furthermore, we note that the Wiener filtering techniques that are discussed in this dissertation and elsewhere in the literature assume that channel (time-varying) statistics are available. We propose a new method that estimates such statistics using the coarse channel estimates obtained through pilot symbols. The effectiveness of the method is shown through simulations.

The thesis is concluded in Chapter 7. Suggestions for continuation of the research presented in this thesis are also presented in this chapter.

# CHAPTER 2

# CHANNEL MODELS AND ESTIMATION

Channel modeling is important in wireless communication because an efficient channel model is essential for the system analysis, design, and development. Moreover, a correct channel model is meaningful for parameter optimization, algorithm testing, and performance evolution of communication systems. Therefore, the first task of the chapter is understanding the channel modeling. Initially, a study of radio propagation is presented, and several important criteria are introduced to characterize wireless channels. Then, we attempt to conduct typical mathematical/statistical channel models that are widely adopted in the literature. The models developed in this chapter will be used in the later chapters of this thesis.

We note that channel esitmation is a vital part of receivers in wireless communication systems. In order to recover the message, the effect of the wireless channel on transmitted information must be estimated. This is often performed based on an approximated underlying model of the radio propagation channel and developing channel estimation algorithms that can precisely track the variation of the channel. We present an overview of the channel esitmation approaches from the literatures in the second part of this chapter.

## 2.1 Radio Propagation

In communication systems, the performance is eventually determined by the medium carrying the message signal. The medium, referred to as the communication *channel*, can be classified into two groups: wired channel and wireless channel. If a solid connection exists between the transmitter and receiver, the channel is called a *wired channel*. When a solid connection does not exist, this connection is called a *wireless channel*. The *wireless* channel may be further categorized by the environment en-

countered, such as indoor, urban, suburban, underwater, etc. The random and severe behavior of the radio propagation puts fundamental limitations to the performance of wireless communication systems. A reliable communication system has to overcome or take advantage of the propagation characteristics subject to the relative radio environment. Therefore, characterization and modeling of the wireless channel is an essential step for system design and it has been studied comprehensively.

In the content of wireless communications, the simplest channel is the classical additive white Gaussian noise (AWGN) where the channel is characterized by a flat gain of unity and an additive statistically independent white Gaussian noise. In realistic wireless channels, three effects of the radio propagation have to be taken into account: *path loss*, *shadowing*, and *multipath fading*. *Path loss* refers to the reduction of the radio frequency (RF) energy through transmission. It decays exponentially with the distance the RF wave travels. *Shadowing* is the attenuation caused by absorption, reflection, scattering, and diffraction raised from the obstacles between the transmitter and receiver. *Multipath fading* is caused by the multiple reflective paths that result in signal spread across time.

In [18], the aforementioned three effects are categorized in two fading types : *large-scale fading* and *small-scale fading*.

- Large-scale fading represents the average signal power attenuation varying over relatively large distance (100-1000 meters) for path loss and distances proportional to the length of the obstructing object (10-100 meters in outdoor environments and less in indoor environments) for shadowing. In general, the large-scale fading is described in terms of a *mean path loss* as a function of transmission distance and a *log-normally distributed variation* due to the shadowing.

- Small-scale fading refers to the dramatic changes in signal amplitude and phase over small variations (as small as half a wavelength) in the spatial separation between a receiver and transmitter. Small-scale fading is often called *Rayleigh fading* because if there is no line-of-sight (LOS) signal component and the number of multiple reflective paths is large, the envelop of the received signal is statistically described by a Rayleigh distribution.

The emphasis of this chapter and the relevant work in the rest of this thesis are related to small-scale fading.

Considering the transmission of a bandpass signal at carrier frequency $f_c$ with complex envelop $s_b(t)$, the mathematical model of the transmitted bandpass signal is given by

$$s(t) = \Re\{s_b(t) \cdot e^{j2\pi f_c t}\}. \tag{2.1}$$

We are interested in the received bandpass signal affected by the multipath fading. First, we consider the case where there is no relative motion in the environment. Assuming each path is associated with a path length $l_i$ and attenuation $a_i$, the received signal $r(t)$ is the superposition of the copies of $s(t)$ through all paths,

$$r(t) = \sum_i a_i \cdot s(t - \frac{l_i}{c}) = \Re\{\sum_i a_i \cdot s_b(t - \frac{l_i}{c}) \cdot e^{j2\pi f_c(t - \frac{l_i}{c})}\}, \tag{2.2}$$

where $c$ is the speed of light. Let $\tau_i = \frac{l_i}{c}$ be the delay of the received copy through the $i$-path. The complex envelope of the received signal is thus

$$r_b(t) = \sum_i a_i \cdot e^{-j2\pi f_c \tau_i} \cdot s_b(t - \tau_i) = \sum_i a_i \cdot e^{-j\psi_i} \cdot s_b(t - \tau_i), \tag{2.3}$$

where $\psi_i = 2\pi \frac{f_c l_i}{c} = 2\pi \frac{l_i}{\lambda}$ is the phase shift due to the time delay imposed by the $i$-path and $\lambda = \frac{f_c}{c}$ is the wavelength corresponding to the carrier frequency $f_c$.

Now, let us consider the effect of the relative movement in the environment. The movement changes the $i$-path length as $\Delta l_i = -v \cos\varphi_i t$, where $v$ is the speed of the moving object and $\varphi_i$ denotes the angle of arrival (AoA) of path $i$ with respect to the direction of the movement. Therefore, the complex envelope of the received signal in (2.3), affected by the movement, is given by

$$\begin{aligned} r_b(t) &= \sum_i a_i \cdot e^{-j2\pi \frac{l+\Delta l_i}{c}} \cdot s_b(t - \frac{l + \Delta l_i}{c}) \\ &= \sum_i a_i \cdot e^{-j\psi_i} \cdot e^{-j2\pi \cos\varphi_i \frac{v}{\lambda} \cdot t} s_b(t - \tau_i + \frac{v \cdot \cos\varphi_i \cdot t}{c}). \end{aligned} \tag{2.4}$$

The delay $\frac{v \cos\varphi_i \cdot t}{c}$ introduced by the movement is relatively small and can be ignored. Furthermore, let us introduce the *Doppler frequency* $f_d = \frac{f_c \cdot v}{c} = \frac{v}{\lambda}$ and the *Doppler shift* $\nu_i = f_d \cos\varphi_i$. Then, (2.4) is simplified to

$$r_b(t) = \sum_i a_i \cdot e^{-j\psi_i} \cdot e^{j2\pi\nu_i t} s_b(t - \tau_i) \tag{2.5}$$

Equation (2.5) can be interpreted as a system with time-variant impulse response

$$h(t,\tau) = \sum_i a_i e^{-j\psi_i} e^{j2\pi\nu_i t} \delta(t-\tau_i) \tag{2.6}$$

where $\delta(i)$ denotes the Dirac delta function, which has $\delta(i) = 1$ when $i = 0$ and $\delta(i) = 1$ when $i \neq 0$.

Eq. (2.5) indicates that the multipath fading manifests itself in three aspects:

- The amplitude attenuation and the phase shift ($A_i = a_i \cdot e^{-j\psi_i}$)
- The delay of the envelope ($\tau_i$)
- The carrier frequency shift ($\nu_i$)

The first two aspects are related to the delay spread, which is the determined topology of environment, and it is often characterized by the *delay spread*.

### 2.1.1 Frequency-Selectivity and Delay Spread

Let us first focus on the effect of the *delay spread* introduced by multipath propagation; i.e., assume $\nu_i = 0$. This reduces (2.5) to

$$r_b(t) = \sum_i A_i \cdot s_b(t-\tau_i) = h(t) * s_b(t). \tag{2.7}$$

where

$$h(t) = \sum_i a_i \cdot e^{-j\psi_i} \cdot \delta(t-\tau_i) = \sum_i A_i \cdot \delta(t-\tau_i). \tag{2.8}$$

This can be viewed as a time-invariant linear system. In frequency domain, the transfer function of the channel is

$$H(f) = \sum_i A_i \cdot e^{-j2\pi f\tau_i}. \tag{2.9}$$

The effect of the time delay is indicated by delay spread. Delay spread, often represented by the maximum excess delay $\tau_{\max}$, equals the difference between the arrival of the first and the last component of the received signal associated with a single transmitted pulse. The effect of delay spread can be characterized by comparing $\tau_{\max}$ with the symbol interval $T_s$. If $\tau_{\max} \ll T_s$, it is said that the channel is *frequency-flat*, since there is no frequency attenuation due to the delay spread. In

this case, all copies of the received signal through different paths arrive within the one symbol interval and they are not resolvable. On the other hand, when $\tau_{\max} > T_s$, the channel is *frequency selective* and suffers from intersymbol interference. As a result, the channel is referred to as *ISI channel*. In the frequency selective/ISI channel, multipath components are resolvable at the receiver and are compensated for through the channel equalizer.

Although the equalizer is desired for the frequency selective channel, the OFDM technique is able to convert the frequency selective channel into parallel independent frequency flat channels if the subcarrier spacing is appropriately selected. Therefore, in the following discussion of this dissertation, we assume the channel is frequency flat, and the channel estimation algorithm is designed to capture the time variation of channel gain for each subcarrier.

### 2.1.2 Time-Selectivity and Doppler Spread

We investigate the effect of Doppler spread due to the relative motion in the environment. To simplify the analysis, we assume the delay spread is relatively small such that it may be ignored, i.e., $s_b(t - \tau_i) \approx s_b(t)$. Hence, the complex envelope of the received signal is expressed as

$$r_b(t) = s_b(t) \cdot \sum_i A_i \cdot e^{-j2\pi\nu_i t} = s_b(t) \cdot h(t). \tag{2.10}$$

Eq.(2.10) implies that the effect of Doppler shift results in a multiplicative function

$$h(t) = \sum_i A_i \cdot e^{-j2\pi t\nu_i}. \tag{2.11}$$

Here, $h(t)$ is often referred to as complex gain (or channel gain). If a single tone signal (at $f_0$) is transmitted through this channel, the received signal would be spread within the interval $[f_0 - f_d, f_0 + f_d]$. This is called *frequency spread*, and it is determined only by the velocity of the motion in the environment. Similar to the discussion on (2.9), the attenuation of the channel due to phase shift is different at different times, i.e., time-varying. This is called the *time-selectivity* of the multipath channel. The product of the Doppler frequency and symbol time $f_d T_s$, called *normalized Doppler frequency*, is a widely accepted indicator of the time-selectivity.

In Table 2.1, we present a few relevant (important) parameters from in LTE [19] and WiMAX [20]. For the simulation in the later chapters, we set $f_d T_s = 0.01 \sim 0.02$ for a relatively fast time-varying channel according to the values listed in this table.

### 2.1.3 Coherence Time and Coherence Bandwidth

As discussed previously, the delay spread in the time domain causes the frequency-selectivity, while the Doppler spread in the frequency domain introduces the time-selectivity. If both kinds of spread are present, the channel is *dual-selective*.

The *Coherence bandwidth* $W_c$ is defined as the frequency bandwidth over which the correlation between two samples of the channel response taken at the same time but different frequencies falls bellow a predefined value. An appropriate value of $W_c$ is given by

$$W_c \simeq \frac{1}{\tau_{\max}}. \tag{2.12}$$

The *Coherence time* $T_c$ is defined as the period of time over which the fading process is correlated (or equivalently, the period of time after which the correlation between two samples of the channel response taken at the same frequency but different time instants drops below a certain predetermined threshold). The coherence time is related to Doppler spread $f_d$ by

$$T_c \simeq \frac{1}{f_d}. \tag{2.13}$$

Comparing the Coherence time $T_c$ to the symbol/frame duration $T$ provides two widely used concepts [21]:

- when $T \ll T_c$, the channel is said to be slow fading and approximated as constant over $T$
- when $T$ is comparable to $T_c$, the channel is said to be fast fading or time-varying over $T$.

**Table 2.1**. The normalized Doppler frequency in LTE and WiMax

| Standard | $T_s$ ($\mu$s) | $f_d$ | $f_d T_s$ |
|---|---|---|---|
| LTE | 66.7 | 5 - 900 | 0.0003 - 0.06 |
| WiMax | 102.4 | 5 - 300 | 0.0005 - 0.03 |

In the latter case, it is necessary to consider the variation of the fading channel from one symbol interval to the next. It is done by taking a specific channel model and its correlation properties depends on the particular propagation environment and the underlying communication scenario.

## 2.2 Channel Modeling

In general, the deterministic modeling of the multipath fading channel requires the path delay $\tau_i$, phase shift $\psi_i$, and Doppler frequency $\nu_i$ associated with each path. This leads to a difficult model to work with. Simpler methods are obtained through statistical description of channels. For example, in the presence of delay spread, the channel $H(f)$ given by (2.9) can be modeled as a Gaussian random process in the frequency domain. In the presence of Doppler spread, the channel $h(t)$ given by (2.11) can be modeled as a Gaussian random process in the time domain. If both kinds of spread are presented, the time-variant transfer function of the channel $H(t, f)$ can be modeled as a Gaussian random process in both time and frequency domains.

### 2.2.1 Rayleigh and Rician Fading

In the absence of a LOS component, the complex channel gain is expressed in polar coordinates, i.e., $h(t) = r \cdot e^{j\theta}$. The channel fading amplitude $r$ can be viewed as a random variable derived from $r = \sqrt{x_1^2 + x_2^2}$, where $x_1$ and $x_2$ are two zero-mean i.i.d. random variables with the variance $\sigma_h^2$. Accordingly, $r$ has a Rayleigh distribution with the PDF [22]

$$p_r(r) = \frac{r}{\sigma_h^2} e^{\frac{-r^2}{2\sigma_h^2}}, \quad (r \geq 0), \tag{2.14}$$

where $\sigma_h^2$ is the variance of the complex Gaussian random process $h(t)$, i.e.,

$$\sigma_h^2 = \frac{1}{2}\mathrm{E}[|h(t)|^2]). \tag{2.15}$$

The phase shift $\theta$ has uniform distribution over $[0, 2\pi)$, i.e.,

$$P_\theta(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi. \tag{2.16}$$

This is called *Rayleigh fading*.

In the presence of the LOS component, the channel fading amplitude $r$ is modeled as $r = \sqrt{(x_1 + A)^2 + x_2^2}$, where i.i.d. random variables $x_1, x_2 \sim N(0, \sigma_h^2)$ and $A \in \mathcal{R}$

is the amplitude of the LOS component. Thus, $r$ has a Rician distribution and the PDF of $r$ is given by

$$p_r(r) = \frac{r}{\sigma_h^2} e^{-\frac{r^2+A^2}{2\sigma_h^2}} \cdot I_0(\frac{A \cdot r}{\sigma_h^2}), \quad r \geq 0, \tag{2.17}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. The Rician distribution is often described in terms of a parameter $K$, which is defined as the ratio between the power of the LOS component and the variance of the multipath component, i.e., $K = \frac{A^2}{2\sigma_h^2}$, or in decibels

$$K|_{dB} = 20 \log \big(\frac{A}{\sqrt{2}\sigma_h}\big). \tag{2.18}$$

$K$ is often referred to as the Rician factor to specify the Rician distribution. Note that Rayleigh distribution is a special case of Rician distribution for $K = 0$.

### 2.2.2 Autocorrelation and Power Spectrum Density

A Gaussian random process is characterized by its mean and autocorrelation function. Equivalently, the same information is carried by the power spectrum of the process. Given a random channel with the impulse response $h(\tau, t)$, where $t$ is the time when the impulse is applied, its autocorrelation is defined as

$$\gamma_h(\tau_1, \tau_2; t, \Delta t) = \mathrm{E}[h^*(\tau_1, t)h(\tau_2, t + \Delta t)], \tag{2.19}$$

Most channels in practice are wide-sense stationary (WSS), where the joint statistics of the channel measured at two different times $t$ and $t + \Delta t$ depend only on the time difference $\Delta t$. Moreover, if the channel has uncorrelated scattering, i.e., the channel responses associated with the multipath component of delay $\tau_1$ and delay $\tau_2$ are uncorrelated, the correlation function in (2.19) can be simplified as

$$\gamma_h(\tau; \Delta t) = \mathrm{E}[h^*(\tau, t)h(\tau, t + \Delta t)], \tag{2.20}$$

where $\gamma_h(\tau; \Delta t)$ is the average output power associated with the channel as a function of the delay $\tau = \tau_1 = \tau_2$. This is called the wide-sense stationary uncorrelated scattering (WSSUS) model.

For $\Delta t = 0$,$\gamma_h(\tau) \triangleq \gamma_h(\tau; 0)$ is defined as the *power delay profile* or *multipath intensity profile* [23]. Given $\gamma_h(\tau)$, the average and root mean square (RMS) delay spread are typically defined as

$$\mu_{\tau_{\max}} = \frac{\int_0^\infty \tau\gamma_h(\tau)d\tau}{\int_0^\infty \gamma_h(\tau)d\tau}, \tag{2.21a}$$

and

$$\sigma_{\tau_{\max}} = \sqrt{\frac{\int_0^\infty (\tau - \mu_{\tau_{\max}})^2\gamma_h(\tau)d\tau}{\int_0^\infty \gamma_h(\tau)d\tau}}. \tag{2.21b}$$

The typical delay profiles specified in 3GPP LTE are summarized in Table 2.2.

In the case of Doppler spread only, $\tau = 0$ in (2.20). Hence, we define

$$\gamma_h(\Delta t) = \mathrm{E}[h^*(t)h(t + \Delta t)]. \tag{2.22}$$

The *power spectral density* (PSD) or *power density spectrum* of the channel Doppler spread is obtained by taking the Fourier transform of $\gamma_h(\Delta t)$

$$S_h(\nu) = \int_{-\infty}^{\infty} \gamma_h(\Delta t)e^{-j2\pi\nu\Delta t}d\Delta t. \tag{2.23}$$

Next, we present the autocorrelation function and the PSD of *Jakes' Model* [24]. Jake's Model is widely used in the literature and standards. It is obtained by taking the following assumption. In the situation where a receiver with an omnidirectional antenna is moving with a speed $v$, and a large number of reflections of the transmitted signal arrive at the receive antenna from all directions with the same probability, one will find that [24]

$$S_h(\nu) = \begin{cases} \frac{2\sigma_0^2}{\pi f_d\sqrt{1-(\nu/f_d)^2)}}, & |\nu| \le f_d, \\ 0, & |\nu| > f_d, \end{cases} \tag{2.24}$$

**Table 2.2**. Delay power profiles of the LTE channel models

| Model | Number of paths | $\sigma_{\tau_{\max}}$ | $\tau_{\max}$ |
|---|---|---|---|
| Extended Pedestrian A (EPA) | 7 | 45 ns | 410 ns |
| Extended Vehicular A (EVA) | 9 | 357 ns | 2.51 $\mu$s |
| Extended Typical Urban (ETU) | 9 | 991 ns | 5 $\mu$s |

where $2\sigma_0^2$ is the total power at the receiving antenna. The spectrum given by (2.24) is often called Jakes' PSD, and sometimes referred to as Clarke's PSD, since it was first derived by Clarke [25].

Taking the inverse Fourier transform of $S_h(\nu)$, the autocorrelation function of Jakes' model is obtained by

$$\gamma_h(\Delta t) = 2\sigma_0^2 J_0(2\pi f_d \Delta t), \tag{2.25}$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, and $f_d$ is the Doppler frequency shift.

In Fig. 2.1, the PSD and the autocorrelation function of Jakes' Model are illustrated for $f_d = 100$ Hz and $2\sigma_0^2 = 1$.

### 2.2.3 MIMO Channels

In MIMO communication systems, the spatial characteristics of the radio channel have significant effect on the system performance. The large MIMO gain is achieved when the spatial correlation between the gains among antenna pairs is low. Thus, appropriate antenna separation or antenna arrays with polarization in orthogonal/near orthogonal orientations are required. In this thesis, it is assumed that there is no spatial correlation. This assumption is valid if the antenna spacing is greater than half a wavelength of the carrier and the scattering environment is rich in multipaths.

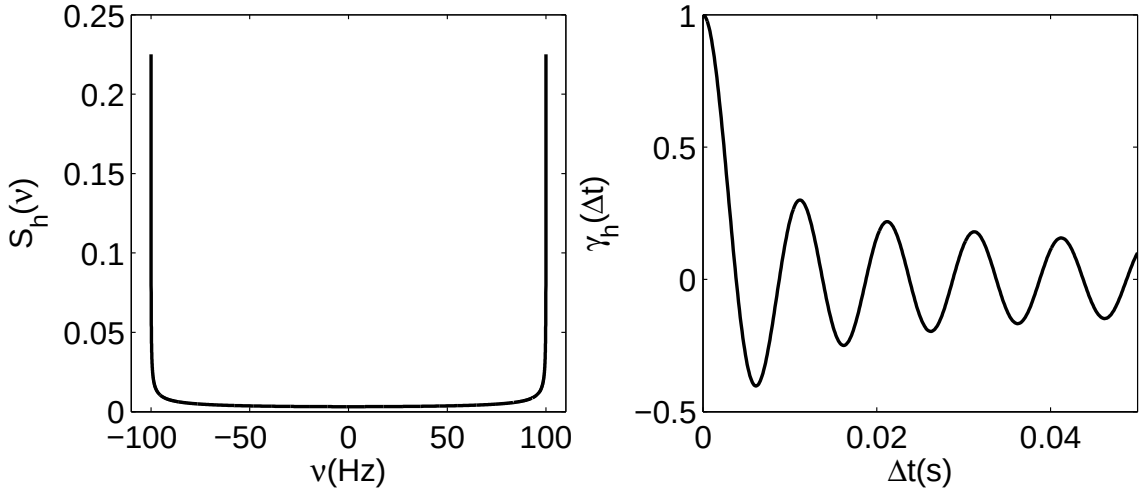


**Figure 2.1**. PSD and autocorrelation function of Jakes' model

An $M \times N$ MIMO system is presented in Fig.1.3. From the system-level perspective, the linear time-variant MIMO channel is represented by an $N$ by $M$ matrix as

$$\mathbf{H}(t,\tau) = \begin{bmatrix} h_{11}(t,\tau) & h_{12}(t,\tau) & \cdots & h_{1M}(t,\tau) \\ h_{21}(t,\tau) & h_{22}(t,\tau) & \cdots & h_{2M}(t,\tau) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1}(t,\tau) & h_{N2}(t,\tau) & \cdots & h_{NM}(t,\tau) \end{bmatrix}, \tag{2.26}$$

where $h_{ij}(t,\tau)$ is the time-variant impulse response of the channel between the $j$-th transmit antenna and the $i$-th receive antenna. In the absence of spatial correlation, the elements $h_{ij}(t,\tau)$ are modeled as a set of statistically independent random processes. To be more specific, each $h_{ij}(t,\tau)$ is an i.i.d complex Gaussian process with zero mean and unit variance.

For a multipath fading channel,

$$h_{ij}(t,\tau) = \sum_{l=1}^{L} a_l(t) e^{-j2\pi f_c \tau_l(t)} \delta(t - \tau_l(t)), \tag{2.27}$$

where $L$ is the number of multipath components, $a_l(t)$ and $\tau_l(t)$ are the complexed-valued attenuation and the excess delay of the $l$-th component, and $f_c$ is the carrier frequency.

As in the case of single links, MIMO channels may also be characterized as frequency-selective or frequency-flat by comparing the maximum delay with the symbol duration. If the maximum delay is small compared to the symbol period, the dependence on $\tau$ vanishes. As a result, the channel is frequency-flat, and the channel impulse response is reduced to a complexed-valued gain. On the other hand, if the channel is time-invariant, the time index $t$ of $h_{ij}(t,\tau)$ is omitted.

However, due to the mobility of the transmitter, receiver, and environmental scatterers, wireless channels are always time varying. To develop channel estimation algorithms, the *quasi-static* assumption is the most common one in the literatures. It assumes that the channel gain is constant in a block of symbols and changes independently in the next block. This is often referred to as *block fading*.

A more realistic channel model that is considered in this thesis is the continuous flat fading channel according to the Jakes' model. As it has been discussed, the time

variation of a flat fading wireless channel can be captured by the normalize Doppler shift $f_d T_s$. The autocorrelation of $h_{ij}$ $(0 \leq j < M,\ 0 \leq i < N)$ is given by

$$\gamma_h(\Delta t) = \mathrm{E}[h_{ij}(t) h_{ij}^*(t + \Delta t)] = J_0(2\pi f_d \Delta t), \tag{2.28}$$

Note that the Doppler frequency shift $f_d$ is assumed to be the same for different antenna pairs.

## 2.3 Channel Estimation

Channel estimation/acquisition is a major challenge for wireless communications employing coherent reception [26]. It becomes more critical in MIMO systems because the promising capacity of MIMO channels relies on the known CSI. Typically, the CSI is estimated by using pilots or training symbols known at both the transmitter and receiver. This technique is often called the *pilot symbol assisted modulation* (PSAM) technique, and it has been widely adopted because of its feasibility and low complexity for implementation. Solid analytical studies on PSAM technique have been provided by [27, 28], etc., for single antenna systems over frequency-flat fading channels. Extended discussions on PSAM for MIMO systems can be found in [29–31]. The optimal training sequence for frequency-flat quasi-static MIMO channels consists of mutual orthogonal pilot symbols [29]. The channel estimation by means of least square (LS) or linear minimum mean-square-error (LMMSE) approaches are studied in [29,31] for quasi-static fading channels, and in [32] for time-varying fading channels. The simple implementation of PSAM technique is at the expense of the reduction in spectral efficiency. However, the spectrum is a limited resource such that spectral efficiency is one of the most important concerns a system designer should account for. Hence, the emphasis of this work is to improve the channel estimation performance constrained to the equal power training sequence with minimum training length.

More powerful channel estimators can be developed by taking advantage of the *turbo principle* and devising methods that detect data and channel iteratively. The *turbo principle* in wireless communications refers to a general approach for combining and serially performing two or more tasks of the receiver in an iterative manner. The idea originated from *turbo codes* first presented in 1993 [12] and has been extended to many fields other than channel coding theory. The quality of the channel estimation

can be improved through iterations by taking advantage of the feedback information from the channel decoder as the uncertain reference signal for channel estimation. In this manner, the channel estimator is referred to as the decision-directed (DD) channel estimator. This method can be further categorized into a hard-DD (HDD) channel estimator and soft-DD (SDD) channel estimator according to the form of the feedback information used for channel estimation.

Iterative channel estimation and data detection for quasi-static MIMO channels has been discussed in [33, 34]. An analytical study on iterative data detection and channel estimation for a frequency-flat quasi-static channel has been developed by Buzzi et al. in [34]. However, this analysis only works for HDD channel estimation with binary phase-shift keying (BPSK) modulation. To the best of our knowledge, much of the available literature on receiver design is limited to the MIMO system using BPSK or quadrature phase-shift keying (QPSK) for data transmission, and claims that the extension to general modulation is straightforward. However, the empirical or theoretical analysis for such an extension is in the absence.

For the time-varying channel, the DD channel estimation algorithm has been proposed in [35] for single antenna systems transmitting BPSK symbols over a frequency-flat time-varying channel. This algorithm was extended to QPSK transmission by Niu and Ritcey in [36]. The important observation in [36] is the fact that for quadrature amplitude modulation (QAM) constellations, the underlying correlation matrix of the channel estimator is data dependent, and thus, its inverse has to be calculated at every channel use. This, clearly, adds significant complexity to the receiver. The turbo receiver and DD channel estimator were extended to MIMO channels in [37–39]. In their works, however, they take the decision values at the channel decoder output as the actual transmitted symbols, i.e., ignoring possible errors in the decisions. This inevitably leads to some loss in performance of the receiver.

The aim of this dissertation is to address a number of key technical challenges in MIMO receiver designs that have not yet been reported. We emphasize joint data detection and channel estimation algorithms for high-rate/large MIMO communication systems. As discussed above, we note that the error propagation in turbo loops is a serious problem that has to be dealt with. In particular, it becomes more pronounced

in large MIMO systems, i.e., MIMO systems with a large number of transmit antennas and/or modulation sizes. The challenges of developing robust joint channel estimation and data detection algorithms for large MIMO systems are addressed in Chapters 5 and 6, for the cases of quasi-static and time-varying channels, respectively.

In Chapter 5, we develop a SDD channel estimation (SDD-CE) that generates a linear MMSE channel estimate based on the soft information of the transmitted symbols obtained from the channel decoder. The proposed SDD-CE provides robust channel estimation for large MIMO systems by taking into account the uncertainty of the data symbols. We also propose a SDD-MCMC detector that combines the SDD-CE with a MCMC detector to achieve near optimal performance for large MIMO systems. Furthermore, based on the SDD-MCMC detector, we develop a decorrelation MCMC detector, termed DEC-MCMC, to further reduce the correlation between the channel estimate and the random samples generated by the Gibbs sampler (GS). Our results demonstrate that DEC-MCMC can better predict the channel estimation error and thus yield superior performance to SDD-MCMC, especially under moderate or fast fading scenarios. In addition, to address the complexity issue of MIMO detectors for large systems, we propose an adaptive MCMC detector, termed ADA-MCMC, to control the detection complexity by adjusting the parameters of the GS according to the channel estimation error at each iteration of the proposed turbo loop. For MIMO systems with 64QAM modulation, our results reveal that ADA-MCMC provides comparable performance to its nonadaptive counterpart with a complexity reduction of 50% or more.

In Chapter 6, we follow the philosophy of [40] and [33] and use soft decisions from the channel decoder to combat the error propagation problem in time-varying channels. We develop the optimal Wiener filter (OWF) for time-varying MIMO channels. However, we note that when the system is large, the complexity of OWF may be unaffordable in practice. Hence, we develop a near optimum channel estimator with a much lower complexity. This method obtains the channel estimate through a two-step procedure that we refer to as *dual-layer*. In the first layer, a set of coarse channel estimates, for all time instants $t$, is obtained through a low complexity process which ignores the channel correlation coefficients. The second layer takes these coarse

estimates and pass them through a smoothing (Wiener) filter that accounts for the channel correlation coefficients. This procedure, which we refer to as the dual-layer Wiener filter (DLWF) channel estimator, is somewhat similar to the channel estimator of [38] and [39]. However, there is a difference. While in [38] and [39] the hard decisions of data symbols from the channel decoder are used to obtain the coarse estimates of the channel, we propose to use soft estimates of the data symbols. Our simulation studies, presented in Chapter 6, show that this modification makes a huge difference in performance.

Another contribution of this thesis is the introduction of MIMO MCMC detectors as an integral part of the developed turbo receiver. Our extensive simulations reveals that, in turbo receivers, the use of the MIMO MCMC detector offers significant gains over the more common detectors, such as soft-MMSE MIMO detectors. In this thesis, we show that part of this gain comes from the fact that the MCMC MIMO detector is more robust to channel estimation than is the soft-MMSE counterpart.

# CHAPTER 3

# MCMC DETECTOR WITH KNOWN CSI

As discussed in Chapter 1, the optimal MIMO detector, known as the maximum likelihood (ML) detector, has a complexity that grows exponentially with the number of bits per channel use. To avoid this complexity, researchers have proposed a number of linear MIMO detectors, such as ZF equalizer, MMSE equalizer, and MMSE equalizer with successive interference cancellation (SIC). Details of these detectors were presented in Chapter 1. These methods reduce the complexity of detectors at the cost of a significant loss in performance. Meanwhile, to achieve near-capacity performance, more elegant detectors were proposed. The list sphere decoding (LSD) [41] and other tree search methods [42] form a class of detectors whose goal is to select a subset of the bit combinations at each channel use as a candidate list that is used for the computation of the LLR values. The candidate list here is obtained through a deterministic approach. Although the size of the list, here, may be significantly smaller than the signal space (the number of all possible bit combinations), it still grows exponentially with the number of bits per channel use [42].

The Markov chain Monte Carlo (MCMC) method [43] is an alternative technique that may also be used to generate a candidate list $\mathcal{L}$ such that $\mathbf{x} \in \mathcal{L}$ yields small values of $\|\mathbf{y} - \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{x}\|^2$ [44–46]. This method is different from the tree-search methods in two ways: (i) it is a stochastic search; (ii) the size of $\mathcal{L}$ does not grow exponentially with the number of bits per channel use. In fact, the complexity of the MCMC-MIMO detector only grows slightly faster than linear.

The emphasis of this chapter is on the mathematical background of the MCMC method and the MCMC that we use for the MIMO detector. The rest of this chapter

is organized as follows. Section 3.1 introduces the *Monte Carlo approximation* and *importance sampling*. Section 3.2 presents the properties of the Markov chain, such as the irreducibility, aperiodicity, and *basic limit theorem* as a fundamental theorem for the MCMC method. Section 3.3 introduces two examples of MCMC method: Metropolis-Hasting algorithm and Gibbs sampling. The latter is the kernel of the original MCMC-MIMO detector which is elaborated in Section 3.4.

## 3.1 Monte Carlo Method

In the context of this thesis, the first "MC" in MCMC stands for the *Monte Carlo* method, which originates from the Monte Carlo approximation in computing an integral and the related important sampling that is discussed below.

### 3.1.1 Monte Carlo Approximation

The problem of the ML detector in (1.13) may be viewed as an optimization problem in the form of

$$\hat{x} = \arg\min_{x\in(a,b)} f(x). \tag{3.1}$$

On the other hand, the MAP detector in (1.19) can be thought of as an integration problem as

$$\theta = \int_a^b g(x)f(x)dx. \tag{3.2}$$

When the underlying functions in (3.1) and (3.2) are simple, the solutions are often obtained analytically. However, when the problem is too complicated to obtain a closed-form solution, Monte Carlo approximation is an alternative option often used. The key idea of Monte Carlo approximation is presented as follows.

For the integration problem in (3.2), if $f(x)$ satisfies

1. $f(x) \geq 0, x \in (a, b)$ and
2. $\int_a^b f(x)dx = C < \infty$,

then $f^*(x) = \frac{f(x)}{C}$ can be viewed as a PDF defined over the interval $(a, b)$. Thus, the integral in (3.2) is equivalent to

$$\theta = \int_a^b Cg(x)f^*(x)dx = C\mathrm{E}_{f^*}[g(x)]. \tag{3.3}$$

The right-hand side of (3.3) is the excepted value of $g(x)$ over distribution $f^*(x)$ scaled by the unknown constant $C$. Monte Carlo approximation suggests drawing a large number of samples $(x_1, x_2, \cdots, x_{N_s})$ from $f^*(x)$ and evaluating $\theta$ as

$$\hat{\theta} = C\frac{1}{N_s}\sum_{i=1}^{N_s} g(x_i). \tag{3.4}$$

By the law of large numbers, it is obvious that $\hat{\theta} \to \theta$ as $N_s \to \infty$ , i.e., the approximation will approach the true value of the integral if a large number of samples is evaluated.

For the optimization problem in (3.1), Monte Carlo approximation is more straightforward since it can be represented by

$$\hat{x} = \arg\max_{x\in(a,b)} f^*(x). \tag{3.5}$$

Given the samples from $f^*(x)$, $\hat{x}$ can be directly estimated by finding the higher density area of $x$ (e.g., plotting the histogram of samples and locating the peak).

### 3.1.2 Importance Sampling

Despite being called importance "sampling", importance sampling has nothing to do with drawing samples. It refers to a MC method of performing the integral in (3.2). To be more specific, importance sampling is a variance reduction technique that is commonly used for Monte Carlo approximation.

Instead of drawing samples from the distribution $f^*(x)$, the samples are alternatively drawn from a proposal distribution $q(x)$ and the integral in (3.2) is computed by the importance sampling method defined as follows.

Given $\{x_i\}_{i=1}^{N_s}$ are i.i.d. samples drawn from the proposal distribution $q(x)$, $\theta = \int_a^b g(x)f(x)dx$ can be approximated by

$$\hat{\theta} = C \cdot \frac{1}{N_s}\sum_{i=1}^{N_s} g(x_i)w(x_i), \tag{3.6}$$

where $w(x) = \frac{f^*(x)}{q(x)}$ is called the weighting factor.

Eq.(3.6) is obtained from applying Monte Carlo approximation to

$$\begin{aligned}\theta &= \int_a^b g(x)\frac{f(x)}{q(x)}q(x)dx \\ &= \int_a^b g(x)\frac{Cf^*(x)}{q(x)}q(x)dx \\ &= C\mathrm{E}_q\{w(x)g(x)\}. \end{aligned} \tag{3.7}$$

Obviously, (3.6) is an unbiased estimate since $\hat{\theta} \to \theta$ as $N_s \to \infty$. Moreover, the variance of *theta* can be calculated as

$$\mathrm{VAR}_{\mathrm{IS}} = \int_a^b g^2(x)\frac{f(x)}{q(x)}f(x)dx - \theta^2. \tag{3.8}$$

Comparing to the variance of Monte Carlo approximation:

$$\mathrm{VAR}_{\mathrm{MC}} = \int_a^b Cg^2(x)f(x)dx - \theta^2, \tag{3.9}$$

we have

$$\mathrm{VAR}_{\mathrm{MC}} - \mathrm{VAR}_{\mathrm{IS}} = C\int_a^b g^2(x)\big[1 - \frac{f^*(x)}{q(x)}\big]f(x)dx. \tag{3.10}$$

This implies that when $q(x)$ satisfies $q(x) \geq f^*(x)$, $\mathrm{VAR}_{\mathrm{MC}} > \mathrm{VAR}_{\mathrm{IS}}$, i.e., the variance from importance sampling is less than the one from Monte Carlo approximation. In particular, if $q(x) = \frac{g(x)f(x)}{\theta}$, (3.8) implies that $\mathrm{VAR}_{\mathrm{IS}} = \theta\int_a^b g(x)f(x)dx - \theta^2 = \theta^2 - \theta^2 = 0$. Hence, by choosing $q(x) = \frac{g(x)f(x)}{\theta}$, the variance of importance sampling is minimized. Although the "optimal" distribution may not be available, since $\theta$ is the unknown quantity to be estimated, it implies that the variance of importance sampling can be significantly reduced by identifying a distribution that is a good approximation to $q(x) = \frac{g(x)f(x)}{\theta}$.

Importance sampling requires the samples can be drawn efficiently from $q(x)$ or form an approximation to $q(x)$. In addition, the weighting factor $w(x)$ has to be evaluated at the sample points $\{x_i\}_{i=1}^{N_s}$. While in most applications, standard importance sampling is not available because the weighting factor $w(x) = \frac{f^*(x)}{q(x)}$ cannot be evaluated in closed-form. In practice, $f(x)$ and $q(x)$ are often known up to a constant, i.e., $\tilde{f}(x) = z_f \cdot f^*(x)$ where $z_f = \int \tilde{f}(x)dx$ and $\tilde{q}(x) = z_q \cdot q(x)$ where

$z_q = \int \tilde{q}(x)dx$. In this manner, the weighting factor of importance sampling can be expressed as

$$\begin{aligned} w(x) &= \frac{f^*(x)}{q(x)} = \frac{\tilde{f}(x)/z_f}{\tilde{q}(x)/z_q} \\ &= \frac{\tilde{f}(x)/\tilde{q}(x)}{\frac{1}{z_q}\int \tilde{f}(x)dx} = \frac{\tilde{f}(x)/\tilde{q}(x)}{\int \tilde{f}(x)/\tilde{q}(x)(\tilde{q}(x)/z_q)dx} \\ &= \frac{\tilde{w}(x)}{\int \tilde{w}(x)q(x)dx}, \end{aligned} \tag{3.11}$$

where $\tilde{w}(x) = \frac{\tilde{f}(x)}{\tilde{q}(x)}$ is called the *normalized weighting factor*, and the integral in the denominator can be evaluated by Monte Carlo approximation as well. Accordingly, the following procedure may be used to evaluate (3.2).

If $\{x_i\}_{i=1}^{N_s}$ are i.i.d. samples draw from a proposal distribution $q(x)$, then $\theta$ can be approximated by

$$\hat{\theta} = C \cdot \frac{1}{N_s}\sum_{i=1}^{N_s} g(x_i)\hat{w}(x_i), \tag{3.12}$$

where $\hat{w}(x_i) = \frac{\tilde{w}(x_i)}{\frac{1}{N_s}\sum_{i=1}^{N_s}\tilde{w}(x_i)}$ is the Monte Carlo approximation of the weighting factor $\tilde{w} = \frac{\tilde{f}(x)}{\tilde{q}(x)}$.

## 3.2 Markov Chain Fundamentals

The key issue while applying Monte Carlo approximation is to draw samples from a desired distribution. MCMC [47] is a method used to draw samples from a proposal distribution by creating a Markov chain. Thus, it is necessary to understand the fundamentals on the Markov chain before we introduce MCMC-MIMO detection.

**Definition 3.1** *Let* $\mathbf{P}$ *be a* $k \times k$ *matrix with elements* $\{P_{i,j} : i, j = 1, \cdots, k\}$. *A random process* $(X_0, X_1, \cdots)$ *with finite state space* $S = \{s_1, \cdots, s_k\}$ *is said to be a* **Markov chain with transition matrix P**, *if for all* $n$, *all* $i, j \in \{1, \cdots, k\}$ *and all* $i_0, \cdots, i_{n_1} \in \{1, \cdots, k\}$ *we have*

$$\begin{aligned} &P(X_{n+1} = s_j | X_0 = s_{i_0}, \cdots, X_n = s_{i_{n-1}}) \\ &\quad = P(X_{n+1} = s_j | X_n = s_i) = P_{i,j}. \end{aligned} \tag{3.13}$$

The elements of transition matrix $\mathbf{P}$ are called transition probabilities.

Let the row vectors $\boldsymbol{\pi}(0), \boldsymbol{\pi}(1), \cdots$ denote the distributions of the Markov chain at times $0, 1, \cdots$, so that

$$\begin{aligned} \boldsymbol{\pi}(n) &= (\pi_1(n), \pi_2(n), \cdots, \pi_k(n)) \\ &= (P(X_n = s_1), P(X_n = s_2), \cdots, P(X_n = s_k)). \end{aligned} \tag{3.14}$$

We call the vector $\boldsymbol{\pi}(0)$ the *initial distribution*. We also note that since $\boldsymbol{\pi}(n)$ represents a probability distribution, we have

$$\sum_{i=1}^{k} \pi_i(n) = 1. \tag{3.15}$$

Given the initial distribution $\boldsymbol{\pi}(0)$ and the transition matrix $\mathbf{P}$, one can find the distributions $\boldsymbol{\pi}(1), \boldsymbol{\pi}(2), \cdots$ of the Markov chain at any time as follows.

**Theorem 3.1** *For a Markov chain $(X_0, X_1, \cdots)$ with state space $S = \{s_1, \cdots, s_k\}$, initial distribution $\boldsymbol{\pi}(0)$, and transition matrix $\mathbf{P}$, for any $n$, the distribution $\boldsymbol{\pi}(n)$ at time $n$ satisfies*

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\mathbf{P}^n. \tag{3.16}$$

For a Markov chain $(X_0, X_1, \cdots)$ with state space $S = \{s_1, \cdots, s_k\}$ and transition matrix $\mathbf{P}$, we say that a state $s_i$ *communicates* with another state $s_j$, writing $s_i \to s_j$, if there exists an $n$ such that

$$(P^n)_{i,j} = P(X_{m+n} = s_j | X_m = s_i) > 0. \tag{3.17}$$

If $s_i \to s_j$ and $s_j \to s_i$, we say states $s_i$ and $s_j$ *intercommunicate*, and write $s_i \leftrightarrow s_j$.

**Definition 3.2** *A Markov chain $(X_0, X_1, \cdots)$ with state space $S = \{s_1, \cdots, s_k\}$ and transition matrix $\mathbf{P}$ is said to be* **irreducible** *if for all $s_i, s_j \in S$, $s_i \leftrightarrow s_j$. Otherwise, the chain is said to be* **reducible**.

An intuitive way to verify that a Markov chain is irreducible is to look at its transition graph, and check that from each state there is a sequence of arrows leading to any other state. Fig. 3.1 shows the examples of irreducible and reducible Markov chains.

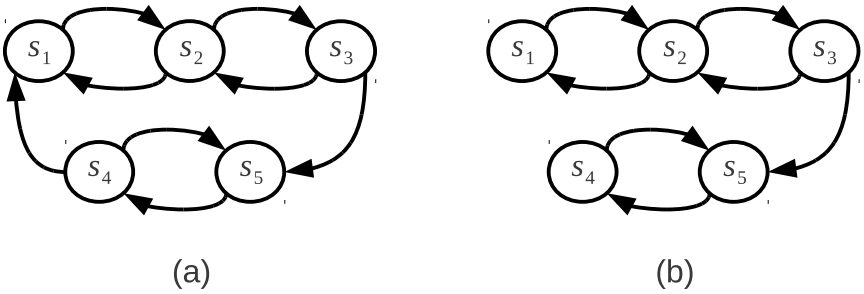


**Figure 3.1**. Transition graphs for (a) an irreducible Markov chain and (b) a reducible Markov chain

**Definition 3.3** *The* **period** *of a state $s_i \in S$ is defined as*

$$d(s_i) = \gcd\{n \geq 1 : (P^n)_{i,i} > 0\}, \tag{3.18}$$

*where* $\gcd\{a_1, a_2, \cdots\}$ *stands for the greatest common divisor of* $a_1, a_2, \cdots$.

Intuitively, a state $s_i$ has a period $d_i$ if any return to state $s_i$ occurs in a time interval that is a multiple of $d_i$. In Fig. 3.2, two examples are given to show the period of a Markov chain.

**Definition 3.4** *An irreducible Markov chain is said to be* **aperiodic** *if the periods of all of its states are* 1. *Otherwise the chain is said to be* **periodic**.

**Definition 3.5** *Let* $(X_0, X_1, \cdots)$ *be a Markov chain with state space* $S = \{s_1, \cdots, s_k\}$

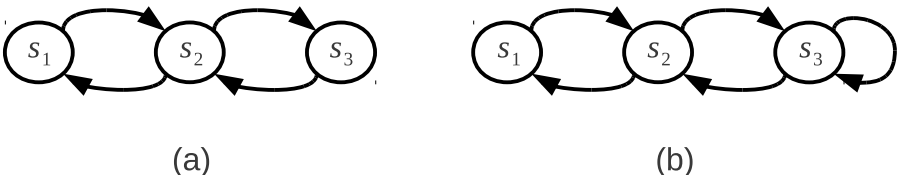


**Figure 3.2**. Transition graphs for Markov chains with (a) period = 2 and (b) period = 1

*and transition matrix* $\mathbf{P}$*. A row vector* $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_k)$ *is said to be a* **stationary distribution** *for the Markov chain, if it satisfies*

*(i)* $\pi_i \geq 0$ *for* $i = 1, \cdots, k$ *and* $\sum_{i=1}^{k} \pi_i = 1$*, and*

*(ii)* $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$*, i.e.,* $\sum_{i=1}^{k} \pi_i P_{i,j} = \pi_j$ *for* $j = 1, \cdots, k$*.*

Finally, we have the following fundamental theorem on the convergence of Markov chains.

**Theorem 3.2** *Let* $X_0, X_1, \cdots$ *be an irreducible aperiodic Markov chain having a stationary distribution* $\boldsymbol{\pi}$*. Then for any initial distribution* $\boldsymbol{\pi}(0)$*,*

$$\lim_{n\to\infty} \boldsymbol{\pi}(0)\mathbf{P}^n = \boldsymbol{\pi}. \tag{3.19}$$

This theorem is called the *basic limit theorem (BLT)*. It claims that if an irreducible aperiodic Markov chain has a stationary distribution, then from any initial distribution $\boldsymbol{\pi}(0)$, the Markov chain will eventually converge to its stationary distribution. For all MCMC methods, how to generate an irreducible and aperiodic Markov chain with a desired stationary distribution is the major question. The next section provides an answer to this question.

## 3.3 Markov Chain Monte Carlo Method (MCMC)

As discussed in Section 3.1, to apply the Monte Carlo approximation to solve the integration problem in (3.4) or the optimization problem in (3.5), a large number of samples have to be drawn from a PDF $f^*(x)$. There are two common algorithms for obtaining sequences of samples from $f^*(x)$.

### 3.3.1 Metropolis-Hasting Algorithm

The Metropolis-Hasting (M-H) algorithm [48–50] draws a sequence of samples of $x$ from the PDF $f^*(x) = \frac{f(x)}{C}$, where $C = \int f(x)dx$ is unknown and is hard to calculate. The M-H algorithm is summarized in Algorithm 3.1.

The M-H algorithm starts with an arbitrary initial value $x = x_0$ that satisfies $f(x_0) > 0$. Then, using the current value $x_0$, a candidate point $x'$ is generated according to the conditional distribution $q(x_2|x_1)$, which is the probability of returning

**Algorithm 3.1:** Algorithm of Metropolis-Hasting

**Input**: $f(x)$, and $q(x_2|x_1)$

**begin**
- Initialization: $x = x_0$;
- **for** $i = 1$ *to* $n$ **do**
  - Draw r.v. $x'$ from $q(x'|x_{i-1})$;
  - Draw r.v. $u$ from $\mathcal{U}(0,1)$, where $\mathcal{U}(a,b)$ stands for uniform distribution on $(a,b)$;
  - Compute $\alpha = \min\left\{1, \frac{f(x')q(x'|x_{i-1})}{f(x_{i-1})q(x_{i-1}|x')}\right\}$;
  - **if** $u < \alpha$ **then**
    - $x_i = x'$ (accept);
  - **else**
    - $x_i = x$ (reject);

**Output**: $x_1, x_2, \cdots, x_n$

a value of $x_2$ given a previous value of $x_1$. This distribution is referred to as *proposal distribution* or *candidate-generating distribution*. The candidate point $x'$ is accepted with the probability defined by

$$\alpha = \min(1, \frac{f(x')q(x'|x)}{f(x)q(x|x')}). \tag{3.20}$$

The above procedure is repeated $n$ times to obtain a Markov chain $\{x_1, \cdots, x_n\}$. Assuming a sufficient *burn-in* process, i.e., that after the first $k$ steps, the Markov chain approaches its stationary distribution, the samples $x_k, x_{k+1}, \cdots, x_{k+n}$ are the samples used for Monte Carlo approximation.

### 3.3.2 Gibbs Sampling

The Gibbs sampling (GS) [51] is a special case of M-H algorithm with the acceptance probability $\alpha = 1$, i.e., the candidate value is always accepted. In addition, GS is efficient for drawing samples from an $m$-dimensional distribution, i.e., the state space consists of vectors, which we refer to as *state vectors*. The key idea of GS is to generate $m$ components of the state vector sequentially from the $m$-univariate conditional distribution rather than from their joint distribution. From this point of view, GS is perfect for the MIMO detector where the state space is spanned by super-symbols consisting of transmitted symbols from all antennas, or multiple bits

**Algorithm 3.2:** Algorithm of GS

**begin**
  Initialization: $\mathbf{x} = \mathbf{x}^{(0)} = (x_1^{(0)}, \cdots, x_m^{(0)})$;
  **for** $i = 1$ *to* $n$ **do**
    **for** $j = 1$ *to* $m$ **do**
      Draw $x_j^{(i)}$ from $f(x_j | x_1^{(i)}, \cdots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \cdots, x_m^{(i-1)})$ ;
    **end**
  **end**
**end**

**Output**: $\mathbf{x}^{(k+1)}, \cdots, \mathbf{x}^{(n)}$

that are transmitted per channel use. The GS algorithm for drawing samples from the joint PDF $f(x_1, x_2, \cdots, x_m)$ is summarized in Algorithm 3.2.

After each round of the inner $j$-loop, a state vector $\mathbf{x}^{(i)}$ is obtained. This is called one *scan*. After $n$ scans, and assuming that the first $k$ scans correspond to the burn in period of the algorithm, the sequence of sample vectors $\mathbf{x}^{(k+1)}, \cdots, \mathbf{x}^{(n)}$ are the desired samples with the joint PDF $f(\mathbf{x})$. The GS algorithm shown in Algorithm 3.2 is the kernel of the MCMC-MIMO detector presented in the next section.

## 3.4 MCMC-MIMO Detector

The goal of the MCMC-MIMO detector is to use GS to generate a set of samples of the transmitted symbol vector $\mathbf{x}$ according to a desired distribution, and then apply Monte Carlo approximation to obtain the estimate (1.19).

### 3.4.1 MCMC-MIMO with Single-GS

In [46], the Monte Carlo Rao-Blackwellization-Uniform (MCRB-U) algorithm is proposed and used as a solution to a low-complexity MCMC-MIMO detector. Next, we introduce the MCRB-U algorithm as an application of the normalized importance sampling that was introduced in Section 3.1.

Recall that the APP value $\mu_1(x_{m,j})$, introduced in Chapter 1, equation (1.19), is calculated as

$$\begin{aligned}\mu_1(x_{m,j}) &\triangleq P(x_m = a_j|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H}) \\ &= \sum_{\mathbf{x}_{-m}\in\mathcal{A}^{M-1}} P(x_m|\mathbf{x}_{-m}, \mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})P(\mathbf{x}_{-m}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H}), \qquad (3.21)\end{aligned}$$

where $\mathbf{x}_{-m} = (x_1, \cdots, x_{m-1}, x_{m+1}, \cdots, x_M)^{\mathrm{T}}$. Obviously, $\mu_1(x_{m,j})$ is the expected value $\mathrm{E}_{f^*(\mathbf{x})}[g(\mathbf{x})]$, where $g(\mathbf{x}) = P(x_m|\mathbf{x}_{-m}, \mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$ and $f(\mathbf{x}) = f^*(\mathbf{x}) = P(\mathbf{x}_{-m}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$. However, it is hard to draw samples from $f^*(\mathbf{x})$ efficiently. Thus, the direct Monte Carlo approximation does not available. To solve this problem, let us make the following assumptions:

(i) $\tilde{f}(\mathbf{x}) = f^*(\mathbf{x})$ ,

(ii) $q(\mathbf{x}) = P(\mathbf{x}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$ , and

(iii) $\tilde{q}(\mathbf{x})$ is a uniform distribution on the space $\mathcal{I}$ that includes the *significant* states of $q(\mathbf{x})$.

It is obvious that $z_f = 1$ from assumption (i). From assumption (ii) and (iii), $\tilde{q}(\mathbf{x}) = \frac{1}{|\mathcal{I}|}$, and the the normalization constant for $\tilde{q}(\mathbf{x})$ is $z_q = \int \tilde{q}(\mathbf{x})d\mathbf{x} = 1$. Therefore, the normalized weighting faction is $\tilde{w}(\mathbf{x}) = |\mathcal{I}|f^*(\mathbf{x}) = |\mathcal{I}|P(\mathbf{x}_{-m}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$. Hence, (1.19) may be approximated by

$$\hat{\mu}_1(x_{m,j}) = \frac{\sum_{i=1}^{N_s} P(x_m^{(i)} = a_j|\mathbf{x}_{-m}^{(i)}, \mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})P(\mathbf{x}_{-m}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})}{\sum_{i=1}^{N_s} P(\mathbf{x}_{-m}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})}, \qquad (3.22)$$

where $\mathbf{x}_m^{(i)} = (x_m^{(i)}, \mathbf{x}_{-m}^{(i)})$ for $i = 0, 1, \cdots N_s$ are i.i.d. samples drawn from the proposal distribution $q(\mathbf{x})$. Although we have $\sum_{i=1}^{N_s} P(\mathbf{x}_{-m}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$ at the denominator, it does not need to be calculated explicitly since it is a constant for all $j = 1, \cdots, 2^{M_c}$. Let $\hat{\mu}_1(x_{m,j}) = \frac{\tilde{\mu}_1(x_{m,j})}{K}$, where $K = \sum_{j=1}^{M_c} \tilde{\mu}_1(x_{m,j})$. By applying the chain rule, $\tilde{\mu}_1(x_{m,j})$ is evaluated as

$$\tilde{\mu}(x_{m,j}) = \frac{1}{P(\mathbf{y}|\boldsymbol{\mu}, \mathbf{H})} \sum_{i=1}^{N_s} P(\mathbf{y}|x_m = a_j, \mathbf{x}_{-m}^{(i)}, \mathbf{H})P(x_m = a_j)P(\mathbf{x}_{-m}^{(i)}), \qquad (3.23)$$

where $\frac{1}{P(\mathbf{y}|\boldsymbol{\mu}, \mathbf{H})}$ is a constant for all $m = 1, \cdots, M$ and $j = 1, \cdots, 2^{M_c}$ such that it can be combined with $K$ to validate $\mu_1(x_{m,j})$ for all $j = 1, \cdots, M_c$ as a PMF.

The sequential MCMC-MIMO (s-MCMC) algorithm runs a single GS to generate samples $\mathbf{x}^{(i)}$, $i = 1, 2, \cdots$, which are i.i.d. according to $q(\mathbf{x})$, and then compute

**Algorithm 3.3:** $s$-MCMC algorithm

**Input**: $n$, $M$, $M_c$, $\mathcal{A}$, $\mathbf{y}$, $\mathbf{H}$, $\boldsymbol{\mu}^e$

**begin**

Initialization: $\mathbf{x} = \mathbf{x}^{(0)}$ and $N_s = 0$;

*%Gibbs Sampling:*

**for** $i = 1$ *to* $I$ **do**

**for** $m = 1$ *to* $M$ **do**

Draw $x_m^{(i)}$ from $p\Big(x_m|\mathbf{x}_{-m}^{(i)}, \mathbf{y}, \boldsymbol{\mu}, \mathbf{H})\Big)$;

**end**

**if** $\mathbf{x}^{(i)} \neq \mathbf{x}^{(i')}$, $\forall\ i' \in [0, I]$ **then**

Accept $\mathbf{x}^{(i)}$;

$N_s = N_s + 1$;

**else**

Reject $\mathbf{x}^{(i)}$;

**end**

**end**

*% Monte Carlo approximation:*

**for** $m = 1$ *to* $M$ **do**

**for** $j = 1$ *to* $2^{M_c}$ **do**

Compute $\lambda_{m,j}$ according to (3.22);

**end**

**end**

**end**

the APP values according to (3.22). The s-MCMC algorithm is summarized in Algorithm 3.3.

## 3.5 The Dilemma of MCMC-MIMO

MCMC suffers from a well-known drawback: it is often difficult to decide when to terminate the algorithm and conclude the convergence on the Markov chain. That is, how can one assure it is safe to stop the GS such that the samples are truly representatives of the Markov chain stationary distribution? Although there is no efficient tool for such decision making, the efficiency of the MCMC algorithm is more important than the convergence rate in practice. When the full convergence of the MCMC algorithm is not affordable, the parameter $N_s$ is chosen to make the algorithm run a certain number of steps, no matter whether it converges or not. In this manner, the algorithms providing better approximation on (1.19) with less number of samples

are highly demanded.

As shown in Algorithm 3.3, the $s$-MCMC algorithm is a combination of the GS and the Monte Carlo approximation in the form of normalized importance sampling. To assess the performance of the MCMC-MIMO algorithm, it is necessary to take the following two aspects into account: the convergence property of the GS and the efficiency/correctness of the Monte Carlo approximation. In fact, both aspects are related to the proposal distribution $q(\mathbf{x})$. First of all, the stationary distribution of the Markov chain should be $q(\mathbf{x})$. Second, $q(\mathbf{x})$ shall be chosen carefully to have a good Monte Carlo approximation. Thus, the proposal distribution $q(\mathbf{x})$ plays a key role in the convergence analysis of the MCMC method.

Recall that the samples to compute $\mu_1(x_{m,j})$ in (3.22) are drawn from a uniform distribution on $\mathcal{I}$ which includes the most significant samples. To be more specific, it is required that the states in $\mathcal{I}$ have relatively large values of $P(\mathbf{x}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$, and are equally likely to occur as well. The uniform hypothesis is supported by the additional "acceptance-rejection" step after the sample $\mathbf{x}^{(i)}$ being drawn in the GS. However, it is not assured that $\mathcal{I}$ includes all significant samples because of the underlying drawback of the GS explained as follows.

To guarantee the samples drawn in GS are from the correct distribution, it is required that the Markov chain be ergodic. A sufficient condition for the ergodicity is that all states in the space be communicable. However, when applying GS to generate multivariate samples, one variable is drawn by conditioning on the other variables. As a result, the successive samples are highly correlated and the Markov chain is likely to keep visiting the same state with larger $P(\mathbf{x}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$. Since $P(\mathbf{x}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$ is proportional to $P(\mathbf{y}|\mathbf{x}^{(i)}, \mathbf{H})P(\mathbf{x}^{(i)})$, the probability $P(\mathbf{x}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$ is relatively large if the SNR is high and/or the detector is overconfident on the prior information. It has been confirmed by the observation of reducing size of $\mathcal{I}$ in high SNR regimes or later iterations. In an extreme case, when $P(\mathbf{x}^{(0)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H})$ is large and $P(\mathbf{x}^{(i)}|\mathbf{y}, \boldsymbol{\mu}^e, \mathbf{H}) \approx 0$, $\forall i \neq 0$, $\mathbf{x}^{(0)}$ will be recurrent all the time and is termed as a *local mode*. When this happens, we say the Markov chain is sticking at a local mode and no longer ergodic. Thus, the samples from the GS are not truly representatives of the desired distribution. As a consequence, the performance of the MCMC-MIMO detector is

degraded, especially in high SNR regimes.

A side effect of the highly correlated samples from the GS is that some significant samples might be missed in $\mathcal{I}$. Therefore, the assumption that $\tilde{q}(\mathbf{x})$ covers the significant region of $q(\mathbf{x})$ will not be satisfied. As a result, the variance of importance sampling will not be reduced as expected. When this happens, more samples (i.e. a larger $N_s$ in (3.22)) are required by importance sampling to obtain an accurate approximation.

When the local mode is triggered, or $\mathcal{I}$ does not include all important samples, the performance of MCMC algorithms are intend to be degraded. In the next chapter, several algorithms are proposed to deal with the potential failure of the MCMC algorithm.

# CHAPTER 4

# IMPROVING THE EFFICIENCY OF MCMC-MIMO DETECTOR

Markov Chain Monte Carlo methods have recently been applied as front-end detectors in multiple-input multiple-output (MIMO) communication systems. Moreover, the near-capacity behavior of such detectors in low signal-to-noise ratio (SNR) regimes have been demonstrated through computer simulations. However, it has also been found that the MCMC-MIMO detectors degrade in high SNR regimes. This chapter investigates into the source of this degradation and proposes a number of ad hoc methods to resolve this undesirable behavior of the MCMC-MIMO detectors. The effectiveness of the proposed methods is shown through empirical (simulation) results.

As noted in earlier chapters, the optimal MIMO detector, known as the maximum likelihood (ML) detector, has a complexity that grows exponentially with the number of bits per channel use. For example, in a MIMO system with 4 transmit antennas, when 4 independent 16-QAM symbols are transmitted from the antennas, the number of bits per channel use is 16. A true MIMO detector has to explore all $2^{16}$ possible combinations of the transmitted bits to extract the required soft information/LLR values. To avoid this complexity, researchers have proposed a number of suboptimal MIMO detectors. Examples are zero forcing (ZF) equalizer [3], minimum mean square error (MMSE) equalizer [52], and MMSE equalizer with successive interference cancellation (SIC) [11]. These methods reduce the complexity of detectors at the cost of significant performance loss. Meanwhile, to achieve near-capacity performance, more elegant detectors were proposed. The list sphere decoding (LSD) [41] and other tree-search methods [42] form a class of detectors whose goal is to select a subset of the bit combinations at each channel use as a candidate list that is used

for the computation of the LLR values. The candidate list here is obtained through a deterministic approach. Although the size of the list, here, may be significantly smaller than the signal space (the number of all possible bit combinations), it still grows exponentially with the number of bits per channel use [42].

The Markov chain Monte Carlo (MCMC) method in [43], also introduced in the previous chapters, is an alternative search technique that may be used to generate a candidate list [44–46]. This method is different from the tree-search methods in two ways: (i) it is a stochastic search; (ii) the growth of the size of the list and thus the complexity of the MIMO detector is not exponential with the number of bits per channel use. In fact, the complexity of the MCMC-MIMO detector only grows slightly faster than the linear. However, the past studies have shown that while the MCMC-MIMO detector performs very well in low SNR (near capacity) regime, it may suffer from an error floor, or even its performance may degrade as SNR increases. The goal of this chapter is to investigate and identify the source of this undesirable behavior of the MCMC-MIMO detector and propose a number of methods that resolve this shortcoming. The content of this chapter was published in [53].

## 4.1 System Model

The block diagram of an $M$-by-$N$ MIMO system is shown in Fig. 4.1. At the transmitter, the information sequence $\mathbf{b}$ is encoded by the channel encoder. The output of the channel encoder after passing through the interleaver is divided into the blocks of $M \cdot M_c$ bits. These blocks form a vector sequence $\mathbf{d}(t)$, where $t$ is the time index. Each $\mathbf{d}(t)$ is then mapped to the transmit symbol $\mathbf{x}(t) = [x_1(t),\ , x_2(t),\ \cdots, x_M(t)]^{\mathrm{T}}$. We assume that each element of $\mathbf{x}(t)$ carries $M_c \cdot M$ coded bits and thus is chosen from a $2^{M_c}$-ary QAM/PSK constellation. We note that each value of $t$ corresponds to one channel use and during each channel use, $M \cdot M_c$ coded bits are being transmitted. In the sequel, since most of our derivations correspond to one channel use, i.e., a fixed $t$, we drop the time index $t$, for brevity.

Assuming a flat fading channel, the received signal can be modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{4.1}$$

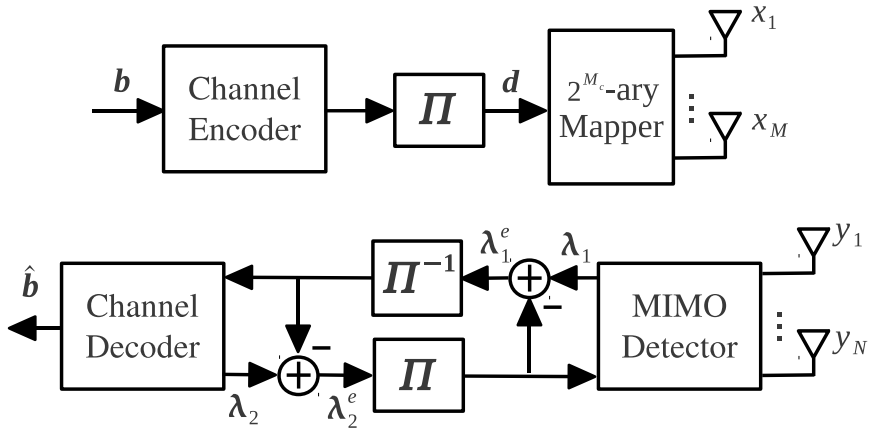


**Figure 4.1**. Block diagram of a MIMO system with soft detector

where $\mathbf{H}$ is the channel gain matrix and $\mathbf{w}$ is the channel noise, a white Gaussian noise vector. We assume that $\mathbf{w}$ has zero mean and the covariance matrix $\mathrm{E}[\mathbf{w}\mathbf{w}^{\dagger}] = \sigma^2\mathbf{I}$. We also note that $\mathbf{x}$ is the transmit vector that is obtained from a block of coded bits represented by $\mathbf{d}$.

## 4.2 Detection Methods

At the receiver, the MIMO detector provides the LLR values

$$\lambda_1(d_k) = \ln \frac{p(d_k = +1|\mathbf{y}, \boldsymbol{\lambda}_2^e(\mathbf{d}))}{p(d_k = -1|\mathbf{y}, \boldsymbol{\lambda}_2^e(\mathbf{d}))}, \tag{4.2}$$

where $d_k$ is the $k$-th element of $\mathbf{d}$, and $\lambda_2^e(\mathbf{d})$ is the extrinsic information from the channel decoder. Here, $\lambda(d_k)$ is defined as the bit-wise LLR. Note that in previous chapters, as well as in Chapter 5 and 6 (to follow), we present the formulations in terms of APPs of data symbols that are denoted by $\mu$. The extrinsic information $\lambda_1^e(d_k) = \lambda_1(d_k) - \lambda_2^e(d_k)$ is then formed and passed to the channel decoder. By exchanging the extrinsic information between the MIMO detector and the channel decoder iteratively, the turbo principle is applied. This procedure reduces the bit-error rate (BER) over successive iterations and allows one to achieve a near capacity performance [41, 46].

Using the max-log approximation, we obtain

$$\lambda_1^e(d_k) \approx \max_{\mathbf{d}\in\mathcal{D}_k^1}\left\{-\frac{\|\mathbf{y}-\mathbf{Hx}\|^2}{2\sigma^2}+\frac{1}{2}\mathbf{d}_{-k}^{\mathrm{T}}\boldsymbol{\lambda}_{2,-k}^{e}\right\} - \max_{\mathbf{d}\in\mathcal{D}_k^0}\left\{-\frac{\|\mathbf{y}-\mathbf{Hx}\|^2}{2\sigma^2}+\frac{1}{2}\mathbf{d}_{-k}^{\mathrm{T}}\boldsymbol{\lambda}_{2,-k}^{e}\right\}, \quad (4.3)$$

where $\mathcal{D}_k^1$ is the set of $\mathbf{d}$ with $d_k = 1$, $\mathcal{D}_k^0$ is the set of $\mathbf{d}$ with $d_k = 0$, $\mathbf{d}_{-k}$ is obtained from $\mathbf{d}$ by removing $d_k$, and $\boldsymbol{\lambda}_{2,-k}^{\mathrm{e}}$ is the vector of the extrinsic LLR values of $\mathbf{d}_{-k}$ from the channel decoder.

The key point and the main reason that has initiated the development of the tree-search methods (including the LSD) and the MCMC-MIMO detector is that the complexity of realization of (4.3) grows exponentially with the number of bits in each channel use. In a MIMO system with $M \cdot M_c$ bits per channel use, each of the sets $\mathcal{D}_k^1$ and $\mathcal{D}_k^0$ have the size of $2^{M\cdot M_c-1}$. Both the tree-search methods and the MCMC-MIMO detector are designed to find small subsets of $\mathcal{D}_k^1$ and $\mathcal{D}_k^0$ that with a high probability contain the desired terms that maximize both terms on the right-hand side of (4.3).

The MCMC-MIMO detector uses a stochastic search method called Gibbs sampler. The Gibbs sampler is a particular Markov chain process that searches the state space defined by $\mathbf{d}$. It walks through this space in a stochastic manner with the goal of finding the samples of $\mathbf{d}$ that result in small values of $\frac{\|\mathbf{y}-\mathbf{Hx}\|^2}{2\sigma^2} - \frac{1}{2}\mathbf{d}_{-k}^{T}\boldsymbol{\lambda}_{2,-k}^{e}$. In other words, the Gibbs sampler looks for important samples of $\mathbf{d}$ that maximize the two terms on the right-hand side of (4.3). For details of the Gibbs sampler, when applied to MIMO detection, we request the reader to refer to Chapter 3 and [46]. Also, for a comparison of the MCMC-MIMO detector and LSD, the reader may refer to [45]. A hardware architecture for efficient implementation of the MCMC-MIMO detector can be found in [54].

## 4.3 MCMC-MIMO Detector in High SNR Regimes

Studies performed in [46] have revealed that while the MCMC-MIMO detector performs very well in low SNR regimes, it does not perform so well as SNR increases. The source of this behavior was explored in Chapter 3. It was noted that at higher values of SNR, some of the transition probabilities in the underlying Markov chain may become very small. As a result, the Markov chain may effectively be divided into

a number of nearly disjoint chains. The term nearly disjoint here means the transition probabilities that allow movement between the disjoint chains are very small. As a result, a Gibbs sampler that is started from a random point will remain within the set of points surrounding the initial point and thus may not get a chance of visiting sufficient points to find the maxima of the terms on the right-hand side of (4.3). In [46], two solutions for solving this problem were proposed: (i) run a number of parallel Gibbs samplers with different starting points; (ii) while running the Gibbs samplers, assume a noise variance higher than what actually is and use the correct noise variance while evaluating (4.3). These two methods turned out to be effective for low and medium size SNRs, as is evident from the excellent results presented in [45, 46, 55].

In many situations in practice, communication systems operate in SNR ranges that are relatively high, many decibels away from the capacity. In such cases, if the MIMO detector can obtain reasonably correct values for the LLRs, one would expect to detect the transmitted information with a very low probability of error through the channel decoder and without any need to run any extra iteration between the MIMO detector and the channel decoder. However, simulations, some of which are presented below, reveal that the above two measures are insufficient to remedy the problem. In this chapter, we propose additional methods to resolve the problem of the MCMC-MIMO detector in high SNR regimes. We also introduce a trivial, yet novel, method for minimizing the receiver complexity.

### 4.3.1 Nonturbo Receiver

We note that, in the absence of the extrinsic information from the channel decoder, the desired solutions that maximize the two terms on the right-hand side of (4.3) are those that result in relatively small values for $\|\mathbf{y} - \mathbf{Hd}\|$. Such solutions are known and can be obtained using a ZF or MMSE equalizer. Through computer simulations, we have found that by initializing one of the Gibbs samplers using either ZF or MMSE solution and initializing the rest of the Gibbs samplers randomly, we obtain results that are much better than those that would be obtained if all of the Gibbs samplers were initialized randomly.

Fig. 4.2 presents a sample of our simulation results. Here, we simulate a 4-by-4

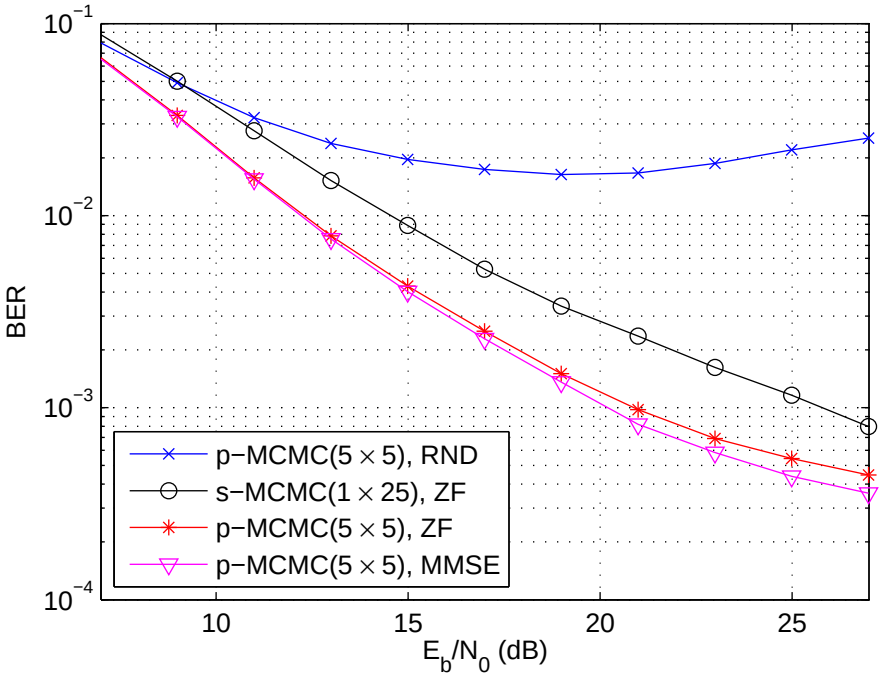


**Figure 4.2**. BER results of a number of different implementations of the MCMC detector

MIMO system. The channel code is the rate $R = 1/2$ convolutional code with the generator polynomials 1 and $1+D^2+D^7$. The data are transmitted in packets of length 1600 uncoded (3200 coded) bits. The channel $\mathbf{H}$ is random, but quasi static, meaning that it is fixed over each packet. However, it is chosen independently for each packet. The elements of $\mathbf{H}$ are complex-valued Gaussian i.i.d. zero mean random variables with variance of unity. Each packet contains a preamble of length 16 that is used for channel estimation, and the estimated channel is used for data detection. There are $M = 16$ bits per channel use. The 16 bits are divided into 4 blocks of 4 bits and mapped to 16-QAM symbols using Gray coding. For the results presented in Fig. 4.2, there is no iteration between the channel decoder and the MIMO detector. The soft information generated by the MIMO detector is passed to the channel decoder and the output of the channel decoder is use to decide on the information bits. We use the normalized SNR [55]

$$\left.\frac{E_b}{N_0}\right|_{\mathrm{dB}} = \left.\frac{E_s}{N_0}\right|_{\mathrm{dB}} + 10\log_{10}\frac{N_r}{N_t R M_c}. \tag{4.4}$$

There are four plots in Fig. 4.2. The first plot is obtained by running 5 parallel randomly initialized Gibbs samplers. Each Gibbs sampler has depth of 5, i.e., it runs over the elements of $\mathbf{d}$ 5 times. This will result in $5 \times 5 = 25$ samples for each of the terms on the right-hand side of (4.3). The second plot is obtained by running a single Gibbs sampler, initialized with the ZF solution, and for a depth of 25; so the sample sets have the same size as in the first plot. The third and fourth plots are generated using 5 parallel Gibbs samplers each of depth 5, with 4 of the Gibbs samplers initialized randomly and the 5th one initialized with the solution obtained from ZF and MMSE equalizers, respectively.

The following conclusions are drawn from the results shown in Fig. 4.2.

- The use of only randomly initialized Gibbs samplers results in a MIMO detector that degrades at high SNR values.
- The use of a single Gibbs sampler initialized with the ZF (or MMSE) solution results in a much improved performance.
- The combination of a number of randomly initialized Gibbs samplers and one Gibbs sampler that is initialized with the ZF (or MMSE) solution further improves the results.
- The level of improvement achieved through ZF and MMSE initializations is about the same.

We have the following explanation to these observations. When Gibbs samplers are initialized randomly, there is always a chance that none of the Gibbs samplers do not approach the portions of the state-space defined by $\mathbf{d}$ that correspond to the maximum terms on the right-hand side of (4.3). As a result, for a relatively large percentage of the channel uses, the MIMO detector may generate incorrect LLR values. The ZF (or MMSE) initialization has a very high likelihood of giving an initial $\mathbf{d}$ within the vicinity of the points that maximize the two terms on the right-hand side of (4.3). The randomized Gibbs samplers result in some level of improvement by adding more samples to the list in the cases where ZF (or MMSE) fails in giving a good initial point. The fact that both ZF and MMSE initialization results in the same

improvement can be explained if we realize in high SNR, where such initializations help, the solutions to both cases are about the same.

From the above results, we observe that although the combination of ZF (or MMSE) and randomized initialization of the Gibbs samplers greatly helps in reducing the BER in high SNR regimes, the BER curves presented in Fig. 4.2 still show some error floor. A number of approaches can be taken to further improve the performance of the receiver. One approach is to increase the number of Gibbs samplers and/or increase their depth. Fig. 4.3 present a sample result that shows how this measure helps. Here, by increasing the number of parallel Gibbs samplers from 5 to 15 and the depth of each Gibbs sampler from 5 to 15 (a 9-fold increase in complexity), we can achieve two orders of magnitude improvement in BER. However, the error floor problem is not resolved.

The following additional measures may be used to improve on the above BER curves and hopefully remove the error floor. (i) Add an additional code with error correcting capability (such as a Reed-Solomon code) prior to the channel encoder.

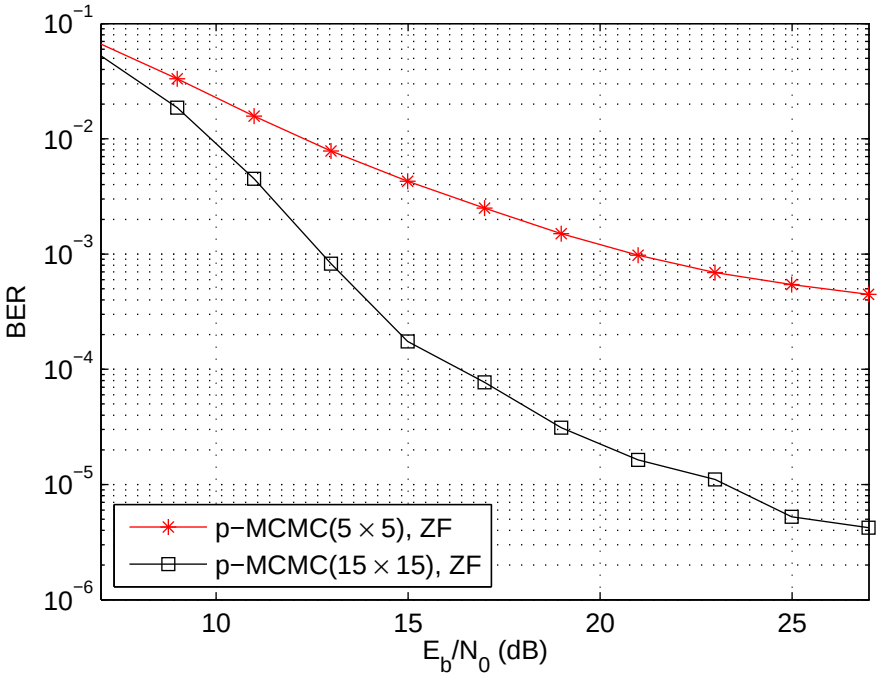


**Figure 4.3**. BER results that show the impact of the number of Gibbs samplers on the receiver performance

The presence of such code can get rid of the residual errors, as long as the number of errors is sufficiently small. (ii) Run iterations between the MIMO detector and the channel decoder. We pursue the latter approach next.

### 4.3.2 Turbo Receiver

To reduce the receiver complexity, we first note when SNR is high and sufficiently accurate estimates of the LLR values are generated by the MIMO detector, error-free recovery of a good majority of the packets occurs in the first iteration of turbo loop. In other words, most of the packets are recovered after the first pass through the MIMO detector and the channel decoder. We thus suggest by adding a parity check (e.g., a CRC check [56]) to each packet, one may examine the correctness of the detected packet. If the packet is detected correctly, no further iteration of the receiver will be executed. If not, soft information from the channel decoder is fed back to the MIMO detector to continue with the second iteration. Similarly, if after the second iteration, the parity check still does not confirm the correctness of the detected packet, iterations continue until the packet is correctly detected or the detection process is terminated after a maximum number of iterations is reached.

Other measures that we empirically (i.e., through computer simulations) found improve the performance of the receiver are:

- After each iteration, one may use the soft information from the channel decoder to randomize the initial settings of the Gibbs samplers for the next iteration.

- Although the latter method greatly helps, for some packets, it does not work, no matter how many iterations of the turbo detector is executed. Detailed exploration of the simulation results reveals that in such cases, the number of bit errors increases with iteration number. In other words, the turbo system can be subject to error propagation. We empirically found a good strategy for solving this problem is to restart the detection process if the turbo loop fails to detect the correct packet after a number of iterations. We refer to each restart of the turbo loop as one stage and number the successive stages as 1, 2, 3, $\cdots$.

- As the receiver proceeds with a new stage, the number of parallel Gibbs samplers and/or the depth of each Gibbs sampler is increased. This, obviously, is done to improve the accuracy of the LLR values generated by the MIMO detector.

The simulation results presented in the next section reveal that the above measures lead to a MIMO receiver in which BER converges to zero as SNR increases.

Unfortunately, any theoretical analysis of the MCMC-MIMO detector turns out to be a very difficult task, and as of today, no such analysis is available. We thus proceed with drawing some conclusions based on numerical studies. The numerical results that are presented in this section are for the $4 \times 4$ MIMO system that was introduced in Section 4.3.1. In addition, to be able to check successful detection of data packets after each iteration of the turbo loop, a length 16 CRC parity checker is added to the coded data bits. For each data packet, the turbo detector is run for three stages; namely, Stage 1, Stage 2, and Stage 3. Stage 1 consists of at most 5 iterations and in each iteration, the Gibbs sampler operates based on 25 samples for each bit; 5 Gibbs samplers, each of depth 5, are run. For Stage 2, the number parallel Gibbs samplers is increased to 10, and their depth is extended to 10. The number of iteration is also increased to 7. In Stage 3, the number parallel Gibbs samplers is increased to 20, and their depth is extended to 20. The number of iteration is increased to 9. The detection process stops when CRC check indicates a correctly detected data packet, or when the three stages of the detection are completed without successful detection of the packet. The simulations are run for 100,000 packets.

Table 4.1 presents the percentages of the successfully detected packets after each iteration of the turbo loop, for $E_b/N_0$ values of 10 to 26 dB. The cumulative percentages of the successfully detected packets are also shown in Fig. 4.4. Referring to the results in this table, the following observations are made:

- Most of the packets are correctly detected within the first stage.

- As SNR increases, the number of iterations required to correctly detect each packet decreases.

**Table 4.1**. Percentages of the successfully detected packets at successive iterations and stages of the turbo loop.

| Iteration No. | $E_b/N_0$, dB | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 |
| 1 | 32.064% | 52.883% | 67.912% | 78.483% | 85.945% | 90.600% | 93.686% | 95.311% | 96.557% |
| 2 | 48.988% | 38.067% | 26.892% | 18.219% | 11.773% | 7.709% | 4.883% | 3.445% | 2.550% |
| 3 | 12.370% | 6.459% | 3.888% | 2.487% | 1.733% | 1.242% | 1.016% | 0.884% | 0.610% |
| 4 | 2.702% | 1.242% | 0.770% | 0.504% | 0.329% | 0.273% | 0.242% | 0.188% | 0.133% |
| 5 | 0.817% | 0.409% | 0.194% | 0.138% | 0.114% | 0.082% | 0.074% | 0.056% | 0.077% |
| End of Stage 1 | 96.941% | 99.060% | 99.656% | 99.831% | 99.894% | 99.906% | 99.901% | 99.884% | 99.927% |
| 6 | 0.042% | 0.021% | 0.000% | 0.011% | 0.003% | 0.003% | 0.003% | 0.000% | 0.000% |
| 7 | 0.188% | 0.151% | 0.056% | 0.040% | 0.013% | 0.005% | 0.005% | 0.011% | 0.005% |
| 8 | 0.456% | 0.204% | 0.111% | 0.048% | 0.040% | 0.035% | 0.035% | 0.050% | 0.024% |
| 9 | 0.332% | 0.093% | 0.053% | 0.027% | 0.032% | 0.032% | 0.016% | 0.032% | 0.019% |
| 10 | 0.127% | 0.045% | 0.011% | 0.016% | 0.013% | 0.013% | 0.021% | 0.016% | 0.014% |
| 11 | 0.053% | 0.013% | 0.005% | 0.003% | 0.003% | 0.000% | 0.011% | 0.000% | 0.007% |
| 12 | 0.027% | 0.000% | 0.000% | 0.000% | 0.000% | 0.003% | 0.003% | 0.003% | 0.003% |
| End of Stage 2 | 98.166% | 99.587% | 99.892% | 99.973% | 99.998% | 99.997% | 99.995% | 99.996% | 99.999% |
| 13 | 0.013% | 0.018% | 0.013% | 0.000% | 0.000% | 0.000% | 0.004% | 0.000% | 0.000% |
| 14 | 0.211% | 0.125% | 0.023% | 0.000% | 0.000% | 0.000% | 0.000% | 0.000% | 0.001% |
| 15 | 0.210% | 0.085% | 0.013% | 0.007% | 0.001% | 0.000% | 0.001% | 0.000% | 0.000% |
| 16 | 0.119% | 0.040% | 0.010% | 0.006% | 0.000% | 0.003% | 0.000% | 0.003% | 0.000% |
| 17 | 0.113% | 0.000% | 0.010% | 0.004% | 0.000% | 0.000% | 0.000% | 0.001% | 0.000% |
| 18 | 0.158% | 0.007% | 0.011% | 0.003% | 0.000% | 0.000% | 0.000% | 0.000% | 0.000% |
| 19 | 0.063% | 0.000% | 0.008% | 0.002% | 0.000% | 0.000% | 0.000% | 0.000% | 0.000% |
| 20 | 0.035% | 0.000% | 0.007% | 0.000% | 0.000% | 0.000% | 0.000% | 0.000% | 0.000% |
| 21 | 0.034% | 0.029% | 0.005% | 0.002% | 0.001% | 0.000% | 0.000% | 0.000% | 0.000% |
| End of Stage 3 | 99.122% | 99.891% | 99.992% | 99.997% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |

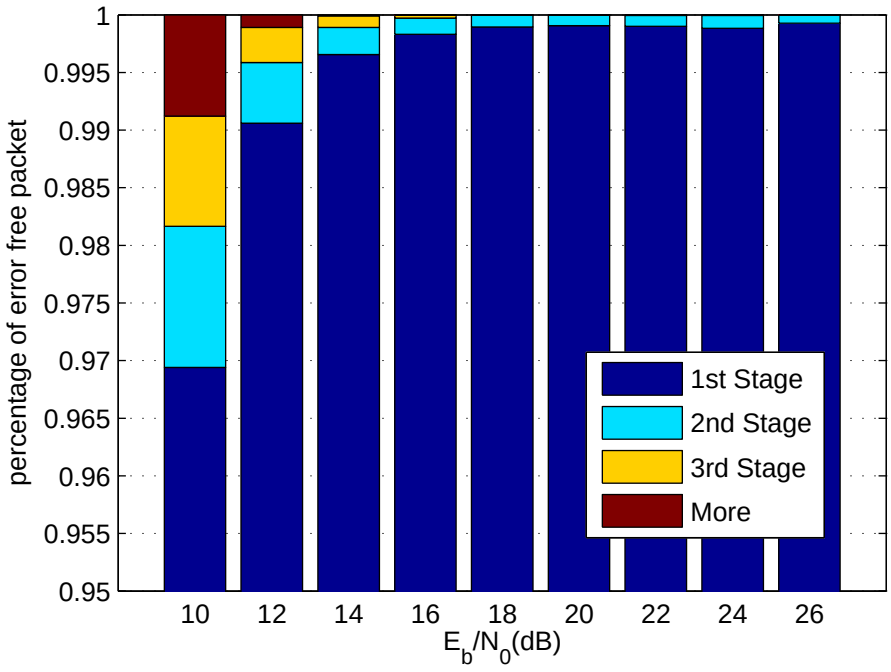


**Figure 4.4**. Cumulative percentages of the successfully detected packets at each stage.

- At higher values of $E_b/N_0$, a large percentage of the packets are correctly detected within the first iteration. For instance, at $E_b/N_0 = 22$ dB, 93.686% of the packets are correctly detected within the first iteration. This number increases to 95.311% at $E_b/N_0 = 24$ dB and to 96.557% at $E_b/N_0 = 26$ dB.

- For values of $E_b/N_0 \geq 18$ dB, all the packets are correctly detected before completion of the third stage.

- At high SNR, since most of the packets are recovered within the first iteration, the average complexity of the receiver is only slightly more than one iteration of the turbo detector. However, the detection of some packets may require a lot more complexity than the average. In other words, the receiver has to deal with the issue of peak complexity. This, in practice, may be dealt with without adding much to the computational power of the receiver, if the communication channel can be tolerable to some data latency.

## 4.4 Conclusion

We proposed a number of measures to overcome the poor performance of the MCMC-MIMO detector in high SNR regimes. The proposed measures/solutions were studied through computer simulations. They were found to be very effective and able to solve the problem. Error-free detection of 100,000 packets, each of length 1600 uncoded information bits, was observed in the $E_b/N_0$ range of 18 to 26 dB.

# CHAPTER 5

# JOINT CHANNEL ESTIMATION AND MCMC DETECTOR FOR QUASI-STATIC RAYLEIGH FADING CHANNELS

This chapter studies the problem of joint data detection and channel estimation for block fading channels, where the channel is fixed over each block of data and varies independently for the next data block. This study, which is mostly of theoretical interest, is common and widely used in the literature. A more realistic channel model, where the channel varies slowly with time, is considered in the next chapter.

## 5.1 Introduction

Joint data detection and channel estimation provides an effective means for improving the receiver performance in wireless transmissions. Such a technique has been investigated extensively for single antenna channels [35, 40, 57] and for multiple antenna channels with a small number of antennas [16, 33, 58] and lower order modulations. The application of such a technique to large antenna systems with higher order modulations, targeted for high-rate communications, has remain largely unexplored due to the following reasons: (1) The complexity of the optimal maximum *a posteriori* (MAP) detector grows exponentially in the number of transmit antennas and modulation size, which necessitates the use of suboptimal detectors with reduced complexity. (2) The receiver performance becomes increasingly sensitive to channel estimation error as the number of antennas and modulation size increases. Existing channel estimation algorithms such as those based on the expectation-maximization algorithm [59–61] become infeasible due to the increased system dimension. Channel estimation algorithms that are robust against data decision errors need to be developed for large MIMO systems.

This chapter aims to address the key technical challenges in the design of joint data detection and channel estimation algorithms for high-rate communications. We are interested in communication scenarios in which multiple transmit antennas, typically three or four, together with higher order modulations, are employed to achieve a desired transmission rate of above 10 bits/channel use. For such systems, MAP detection becomes prohibitive due to its exponential complexity with respect to system dimension. A central part of our design for low-complexity data detection is based on the Markov Chain Monte Carlo (MCMC) method. In recent years, Markov Chain Monte Carlo (MCMC) detection has emerged as an attractive statistical detection method for multiple-input multiple-output (MIMO) channels. As discussed in Chapter 3 and 4, the core of the MCMC detector is the Gibbs sampler (GS), which is a statistical procedure used to generate random samples of the transmitted signals. The MCMC detector adopts the GS to search for a small (to keep the complexity low) but important (to achieve good performance) sample set that contains the likely transmitted signal vectors. The MCMC-MIMO detector can achieve near-optimal performance of the MAP detector with a substantially reduced complexity that is linear in the number of transmit/receive antennas [45, 46, 53, 62, 63]. It outperforms the linear MMSE detector [45], and also the tree-search-based sphere decoding detectors [45, 63] with a complexity that is orders of magnitude less. While most of the prior work on MCMC detection assumes that the channel state information (CSI) is perfectly known at the receiver [45, 46, 53, 63], in this work, we investigate MCMC detection for large MIMO channels with unknown CSI and show that MCMC detection is a high-performance and low-complexity detection method for joint data detection and channel estimation.

We consider an iterative receiver in which soft information, in terms of the likelihood of the transmitted symbols, are interchanged iteratively between the MCMC detector, channel estimator, and a soft-input soft-output channel decoder. At each iteration, joint MCMC detection and channel estimation is performed based on the soft information from the channel decoder. A key issue that often arises in the framework of joint data detection and channel estimation is the error propagation due to developing channel estimation based on erroneous data decision, and then per-

forming data detection based on biased channel estimation. This issue becomes more pronounced with increasing number of transmit antennas and the modulation size because higher order modulations are more sensitive to channel estimation error. To address this design challenge, the proposed receiver adopts the soft-decision-directed channel estimation (SCE), which generates a linear MMSE channel estimate based on the soft information about the transmitted symbols. The SCE provides robust channel estimation for large MIMO systems by taking into account the uncertainty of the data symbols.

The main contributions of this work are summarized as follows:

1. We propose a SCE-MCMC detector that combines the SCE with MCMC detection to achieve near-optimal performance for large MIMO systems with higher order modulations. Closed-form expressions of SCE and its channel estimation error are derived for arbitrary modulation sizes. A new GS is designed for the MCMC detector to generate random samples in accordance to SCE.

2. Based on SCE-MCMC, we develop a decorrelation MCMC detector, termed DEC-MCMC, to further reduce the correlation between the channel estimate and the random samples generated by the GS. Our results demonstrate that DEC-MCMC can better predict the channel estimation error and thus yield superior performance to SCE-MCMC, especially under moderate or fast fading scenarios.

3. We propose an adaptive MCMC detector, termed ADA-MCMC, to control the detection complexity by adjusting the parameters of the GS according to the channel estimation error at each iteration of joint data detection and channel estimation. For MIMO systems with 64QAM modulation, our results show that ADA-MCMC provides comparable performance to its nonadaptive counterpart with a complexity reduction of 50% or more.

4. For various channels examined in this work, the proposed DEC-MCMC demonstrates substantial performance gain over DEC-MMSE, an improved version of the state-of-the-art soft MMSE detector [15] that we derive following the design principle of DEC-MCMC.

MCMC detection for channels with imperfect CSI has been studied previously in the literature. In [62], a noncoherent MCMC-MIMO detector is proposed under the assumption that only the statistics of the CSI is available at the receiver. While the noncoherent approach is theoretically optimal, it is difficult to extend it to systems with more than two transmit antennas due to the slow convergence of the noncoherent GS. In [64], a list-based MCMC-MIMO detector is designed for joint data detection and channel estimation, in which the GS generates likely pairs of data samples and channel estimates. This approach, however, is prone to error propagation as the system dimension increases. MCMC detection for unknown frequency-selective fading channels has been studied in [65–68]. As opposed to the MCMC detectors of [65–67], the proposed MCMC detector in this work does not require a burning period or utilize bit-counting for computing a posteriori probabilities. This work also differs from the list-based approach in [68] that assumes a single transmit antenna and adopts adaptive channel estimation algorithms to facilitate joint data detection and channel estimation. The proposed design uses SCE, which differs from the Wiener filtering-based approach [35, 57] in that it does not require prior knowledge on the time correlation of the fading process.

The remainder of the chapter is organized as follows. In Section 5.2, we introduce the system model and the principle of the iterative receiver. In Section 5.3, we investigate the robustness of the MMSE detector and the MCMC-MIMO detector to imperfect CSI. In Section 5.4, we present the proposed MCMC detection algorithms in conjunction with SCE. In Section 5.5, we introduce the genie-aided channel estimation MCMC as a more realistic performance benchmark for the proposed design. Simulation results are provided in Section 5.6. Conclusions are given in Section 5.7.

## 5.2 System Model

### 5.2.1 MIMO Channel Model and Transmitter Structure

We consider a MIMO block fading channel with M transmit antennas and N receive antennas that operates in a Rayleigh flat-fading environment. The channel between each pair of transmit and receive antenna is assumed to remain constant for a block of $T$ symbol periods, where $T$ is the coherence time, and is independent from block to

block. The channels between different antenna pairs are assumed to be statistically independent. A block diagram of the transmitter structure of the proposed system can be found in Fig. 5.1.

An information bit sequence $\mathbf{b}$ is encoded by a channel encoder of rate $R$. The coded bit sequence $\mathbf{d}$ is interleaved according to a permutation function $\Pi(\cdot)$ and then mapped to a symbol sequence using a $2^{M_c}$-ary constellation. After inserting pilot symbols, the resulting symbol sequence is mapped to a sequence of signal matrices $\mathbf{X} \in \mathcal{C}^{M\times T}$ and sent through the MIMO block fading channel. Here, the first $M$ columns of $\mathbf{X}$ consists of pilot symbols, and the rest $T-M$ columns consists of data symbols. An average power constraint is imposed such that $\frac{1}{M}\mathrm{E}[\mathbf{x}_t^\dagger \mathbf{x}_t] = 1$, where $\mathbf{x}_t$, for all $t \in \{1, \cdots, T\}$ denotes the $t$-th column of $\mathbf{X}$.

A discrete-time model for the received signal over a block fading channel of coherence length $T$ is given by

$$\mathbf{Y} = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{X} + \mathbf{W}, \tag{5.1}$$

where $\mathbf{Y} \in \mathcal{C}^{N\times T}$ is the received signal matrix; $\rho$ is the signal-to-noise ratio (SNR) at each receive antenna; $\mathbf{H} \in \mathcal{C}^{N\times M}$ is the channel matrix and each matrix element is modeled as an i.i.d. complex Gaussian random variable with zero mean and unit variance; $\mathbf{W} \in \mathcal{C}^{N\times T}$ is the white complex Gaussian noise with zero mean and unit variance.

### 5.2.2 Receiver Structure

The iterative receiver considered in this chapter is shown in Fig. 5.2. It consists of three operating modules, MCMC detector, channel estimator, and channel decoder.

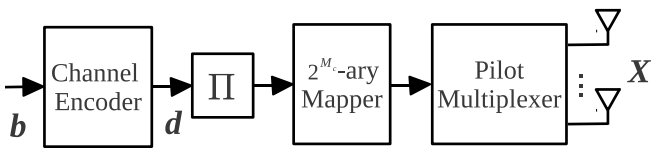


**Figure 5.1**. The block diagram of a coded MIMO transmitter

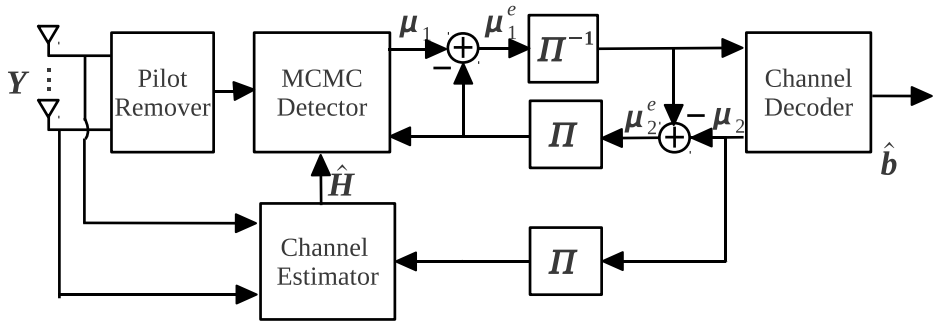


**Figure 5.2**. The block diagram of an iterative MIMO receiver

Soft information, in terms of the probabilities of the transmitted symbols, are interchanged between these modules. We let $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ denote the output *a posteriori probabilities* (APPs) of the transmitted symbols produced by the MCMC detector, and by the channel decoder, respectively. The corresponding extrinsic information (after removing the *a priori probabilities*) is denoted by $\boldsymbol{\mu}_1^e$ and $\boldsymbol{\mu}_2^e$, respectively. At each iteration, given $\boldsymbol{\mu}_2$ and $\mathbf{Y}$, channel estimation is performed to obtain the estimated channel $\hat{\mathbf{H}}$. Subsequently, $\hat{\mathbf{H}}$ and $\boldsymbol{\mu}_2^e$ are fed to the MCMC detector for data detection. The MCMC detector generates updated symbol probabilities $\boldsymbol{\mu}_1$, and the extrinsic information $\boldsymbol{\mu}_1^e$ is passed back to the channel decoder for data decoding. In this way, joint MCMC detection, channel estimation, and data decoding is performed iteratively. After a predetermined number of iterations, decisions are made at the receiver output to obtain the estimated information bit sequence $\hat{\mathbf{b}}$. Details of joint MCMC detection and SCE are provided in Section 5.4.

## 5.3 Robustness to Imperfect CSI

To investigate the impact of channel estimation on the overall system performance, two different MIMO detectors are compared: the soft-MMSE MIMO detector presented in Chapter 1 and the MCMC-MIMO detector discussed in Chapter 3 and 4. To the best of our knowledge, the literature on joint channel estimation and MIMO detection are mostly limited to small systems, for instance, a $2 \times 2$ MIMO system with BPSK modulation as in [33, 35, 38, 39]. However, the interest of this dissertation is large MIMO systems for high-rate communication. The most relevant results to

large MIMO systems are presented in [16]. To achieve a favorable channel estimation, the training transmission power level is boosted 2.5 dB over the payload part in [16]. Considering the amplifier efficiency and the power efficiency, in this thesis, we assume equal power for both training and payload.

To evaluate the system performance with estimated channel gains without specifying any particular channel estimation algorithms, we take the following procedure. Let estimated channel gain $\hat{h}_{n,m}$ be simulated as $h_{n,m} + \epsilon_{m,n}$ where $h_{n,m}$ is the actual channel gain, and $\epsilon_{m,n}$ is a random variable with zero mean and the variance $\sigma^2_{\hat{h}}$. In this manner, the channel estimates are unbiased and the MSE $\sigma^2_{\hat{h}}$ indicates the quality of channel estimation. Then, the receiver employs the imperfect CSI $\hat{h}_{n,m}$ for MIMO detection.

In Fig. 5.3, we compare three coded-MIMO systems: $2 \times 2$ MIMO system with BPSK modulation (small), $3 \times 3$ MIMO system with 16QAM modulation (medium), and $4 \times 4$ MIMO system with 64QAM modulation (large). A rate $R = 1/2$ convolutional code of length 8000 with generator polynomial $(117, 155)_{octal}$ is used. The block length is $T = 10$. Fig. 5.3 shows the curves of the BER versus $\sigma^2_{\hat{h}}$. The operating SNR values shown in Fig. 5.3 are selected to be 5.5dB, 8dB, and 12.5dB for the small, medium, and large systems, respectively. Thus, decent BERs (around $10^{-3} \sim 10^{-4}$) are obtained with perfect channel estimates (i.e., $\sigma^2_{\hat{h}} = 0$). Both the soft-MMSE detector and the MCMC detector are considered. The parameters for the MCMC detector are $G = I = 4$ for the small MIMO system and $G = I = 10$ for the other two systems. We have the following observations.

- The BER curve experiences a *flat-top* as $\sigma^2_{\hat{h}}$ increases. It is because the poor channel estimation would cause severe error propagation such that the system is not able to benefit from any turbo gain. The value of $\sigma^2_{\hat{h}}$, which the flat-top begins with, indicates the system tolerance to channel estimation error. The smaller the value is, the less error-tolerant the system will be. We observed that the flat-top of the small system does not appear even at $\sigma^2_{\hat{h}} = 0.3$, and the BER is around $3 \times 10^{-2}$. It implies that for small MIMO systems, PSAM is sufficient to obtain a good performance. However, the flat-top appears around $\sigma^2_{\hat{h}} = 0.1$

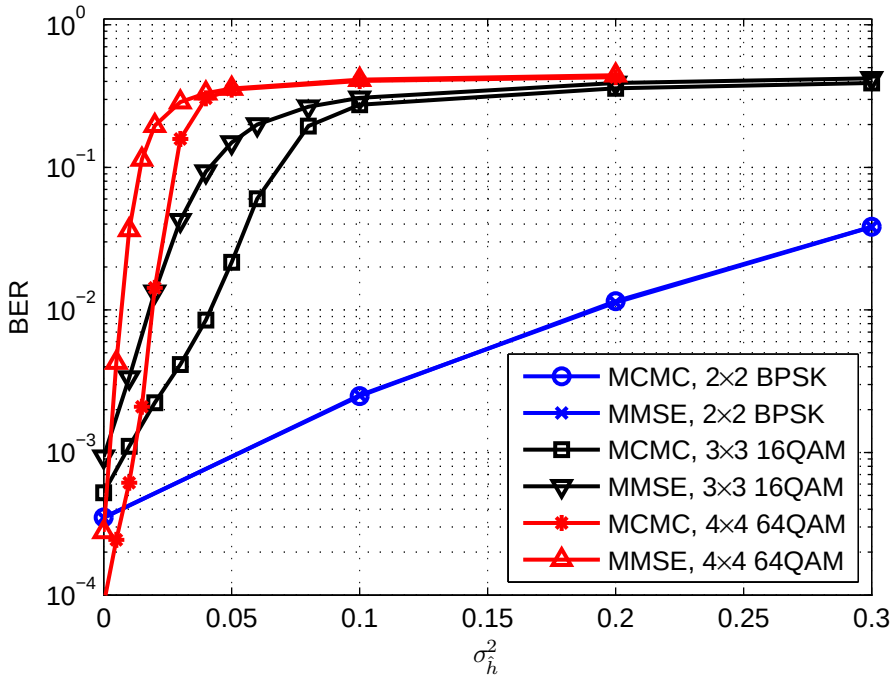


**Figure 5.3**. Sensitivity of MIMO systems to imperfect CSI

for the medium system and $\sigma^2_{\hat{h}} = 0.04$ for the large system., and the resulting BER is around $3 \times 10^{-1}$.

- The *slope* of the BER curve implies how robust the system is to the channel esitmation error. A smaller slope implies the system is more robust to channel estimation error. Obviously, the large MIMO systems shown in the figure are the least robust. For instance, as $\sigma^2_{\hat{h}}$ increases from 0.01 to 0.02, the BER of the $3 \times 3$ MIMO system with 16QAM raises from 0.001 to 0.002 , while that of the $4 \times 4$ MIMO system with 64QAM raises from $6 \times 10^{-4}$ to $1.43 \times 10^{-2}$.
- Given the same channel estimates, i.e., the same value of $\sigma^2_{\hat{h}}$, the MCMC detector outperforms the soft-MMSE detector. Furthermore, the performance degradation caused by the soft-MMSE detector with estimated channel is worse in large MIMO systems. It demonstrates that the MCMC detector is more favorable than the soft-MMSE detector when CSI is not available.

From the above observations, we conclude that the MCMC detector is more robust to channel estimation error than the soft-MMSE detector, and the robustness decreases as system size increases.

## 5.4 MCMC Detection with SCE

In this section, we describe the proposed MCMC detectors based on SCE. Several versions of the MCMC detectors are developed, all of which, in conjunction with SCE, generalize the idealized MCMC-MIMO detector described in Chapter 3 and 4 to provide robust detection performance under the challenging scenario of channel uncertainty. We first present the SCE-MCMC detector in which the MCMC detector is designed to exploit SCE and to take the channel estimation error into account in the detection process. Closed-form expressions of the SCE and its estimation error are derived for general constellation sizes. Based on SCE-MCMC, we develop two advanced versions of MCMC detectors, termed the decorrelation MCMC (DEC-MCMC), aiming to break the correlation between channel estimation and data decisions, and the adaptive MCMC (ADA-MCMC), aiming to control the complexity of the MCMC detector in accordance to the quality of the channel estimation, in Sections 5.4.2 and 5.4.3, respectively.

### 5.4.1 SCE-based Markov Chain Monte Carlo Detection (SCE-MCMC)

The SCE-MCMC operates in two steps. First, we compute the SCE, denoted by $\hat{\mathbf{H}}$, based on the entire block of received signal $\mathbf{Y}$ and the APP $\boldsymbol{\mu}_2$ of the symbols in $\mathbf{X}$. This step takes advantage of the fact the channel remains the same within each block. Second, the MCMC detector performs symbol detection over each column of $\mathbf{X}$. Thus, MCMC detection over different columns of $\mathbf{X}$ is conducted separately once $\hat{\mathbf{H}}$ is available. This design provides superior performance because SCE yields a reliable channel estimate by effectively exploiting the APPs of all symbols in $\mathbf{X}$, and that enables the MCMC detector to successfully separate the interfering symbols transmitted from the $M$ transmit antennas simultaneously.

#### 5.4.1.1 SCE for General Modulations

SCE has been studied in prior work [33, 34, 57] for small modulation sizes such as BPSK. For the completeness of this work, in this section, we present the derivation of SCE for general modulations, following similar techniques in [33].

The SCE estimates each row of channel matrix $\mathbf{H}$ separately. For each $n = 1, \cdots, N$, let $\mathbf{h}_n^r$ denote the $n$-th row of $\mathbf{H}$. The received signal at the the $n$-th receive

antenna, denoted by $\mathbf{y}_n^r$, can be written as

$$\mathbf{y}_n^r = \sqrt{\frac{\rho}{M}}\mathbf{h}_n^r\mathbf{X} + \mathbf{w}_n^r. \tag{5.2}$$

The SCE is obtained by computing the linear minimum mean square error estimator (LMMSE) of $\mathbf{h}_n^r$ based on $\boldsymbol{\mu}_2$. The mean and variance of each transmitted symbol $x_{m,t}$ for all $1 \leq m \leq M$ and $T_p < t \leq T$, denoted by $\bar{x}_{m,t}$ and $v_{m,t}$, respectively, can be computed according to (1.26). The LMMSE estimate $\hat{\mathbf{h}}_n^r$ is derived as

$$\begin{aligned} \hat{\mathbf{h}}_n^r &= \mathbf{y}_n^r \mathrm{E}\big[\mathbf{y}_n^{r\,\dagger}\mathbf{y}_n^r\big]^{-1} E\big[\mathbf{y}_n^{r\,\dagger}\mathbf{h}_n^r\big] &(5.3)\\ &= \sqrt{\frac{\rho}{M}}\mathbf{y}_n^r\boldsymbol{\Gamma}^{-1}\bar{\mathbf{X}}^\dagger\big(\mathbf{I}_M + \frac{\rho}{M}\bar{\mathbf{X}}\boldsymbol{\Gamma}^{-1}\bar{\mathbf{X}}^\dagger\big)^{-1}, &(5.4)\end{aligned}$$

where

$$\boldsymbol{\Gamma} = \mathbf{I}_T + \frac{\rho}{M}\mathrm{E}\big[\big(\mathbf{X}-\bar{\mathbf{X}}\big)^\dagger\big(\mathbf{X}-\bar{\mathbf{X}}\big)\big] \tag{5.5}$$

depends on the second order statistics of the decision error $\mathbf{X} - \bar{\mathbf{X}}$ and the SNR $\frac{\rho}{M}$.

It can be shown that the diagonal elements of $\boldsymbol{\Gamma}$ are $\boldsymbol{\Gamma}(i,i) = 1 + \frac{\rho}{M}\sum_{m=1}^{M} v_{m,i}$ for all $i \in \{1, \cdots, T\}$. The nondiagonal elements of $\boldsymbol{\Gamma}$, as shown in [33], are approximately zero under the assumption that decision errors of the symbol vectors transmitted at different time instances are independent. Thus, we can approximate $\boldsymbol{\Gamma}$ as a diagonal matrix and $\boldsymbol{\Gamma}^{-1}$ in (5.4) can be easily computed to obtain the LMMSE channel estimate. After we apply (5.4) to compute $\hat{\mathbf{h}}_n^r$ for all $1 \leq n \leq N$, we stack these vectors up to form the estimated channel matrix $\hat{\mathbf{H}}$.

#### 5.4.1.2 MCMC Detection Based on SCE

In this section, we describe how to design a new GS for MCMC detection that takes the SCE and its channel estimation error into account. The SCE $\hat{\mathbf{H}}$ is fed to the MCMC detector for data detection over each column of symbols in $\mathbf{X}$. In the following, we drop the time index $t$, and let $\mathbf{y}$ denote the received signal at all receive antennas at a given time instance. Let $\boldsymbol{\Xi} = \mathbf{H} - \hat{\mathbf{H}}$ be the channel estimation error matrix. We can express $\mathbf{y}$ as

$$\mathbf{y} = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{x} + \mathbf{w} = \sqrt{\frac{\rho}{M}}(\hat{\mathbf{H}} + \boldsymbol{\Xi})\mathbf{x} + \mathbf{y} = \sqrt{\frac{\rho}{M}}\hat{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{w}}, \tag{5.6}$$

where $\tilde{\mathbf{w}} = \sqrt{\frac{\rho}{M}}\boldsymbol{\Xi}\mathbf{x} + \mathbf{w}$ denotes the effective noise that takes into account both channel noise and the channel estimation error. Eq.(5.6) is the channel model based

on which the MCMC detector, and the respective GS, has to be built. To use this channel model, we need to evaluate the statistics of the effective noise, $\tilde{\mathbf{w}}$.

Let $\boldsymbol{\xi}_n^r = \mathbf{h}_n^r - \hat{\mathbf{h}}_n^r$ be the $n$-th row of $\boldsymbol{\xi}$. Assuming that there is no spatial correlation between each antenna pair, we have $\mathrm{E}[\boldsymbol{\xi}_n^r \boldsymbol{\xi}_m^{r\,\dagger}] = 0$ for every $n \neq m$. The covariance matrix of the effective noise $\tilde{\mathbf{w}}$ can be computed as

$$\mathbf{C}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}(m,m) = 1 + \frac{\rho}{M}\mathrm{E}\big[\boldsymbol{\xi}_m^r \mathbf{x}\mathbf{x}^\dagger \boldsymbol{\xi}_m^{r\,\dagger}\big] = 1 + \frac{\rho \cdot \|\mathbf{x}\|}{M}\mathrm{E}\big[\boldsymbol{\xi}_m^r \boldsymbol{\xi}_m^{r\,\dagger}\big], \tag{5.7a}$$

and

$$\mathbf{C}_{\tilde{\mathbf{w}}\tilde{\mathbf{w}}}(m,n) = \frac{\rho \cdot \|\mathbf{x}\|}{M} E\big[\boldsymbol{\xi}_m^r \boldsymbol{\xi}_n^{r\,\dagger}\big] \approx 0. \tag{5.7b}$$

To compute $\mathrm{E}\big[\boldsymbol{\xi}_n^r \boldsymbol{\xi}_n^{r\,\dagger}\big]$, we have

$$\begin{aligned} E\big[\boldsymbol{\xi}_n^r \boldsymbol{\xi}_n^{r\,\dagger}\big] &= \mathrm{Tr}\Big\{\mathrm{E}\big[\boldsymbol{\xi}_n^{r\,\dagger} \boldsymbol{\xi}_n^r\big]\Big\} \\ &= \mathrm{Tr}\Big\{\mathrm{E}\big[\mathbf{h}_n^{r\,\dagger}\mathbf{h}_n^r\big] - \mathrm{E}\big[\mathbf{h}_n^{r\,\dagger}\mathbf{y}_n^r\big]\mathrm{E}\big[\mathbf{y}_n^{r\,\dagger}\mathbf{y}_n^r\big]^{-1}\mathrm{E}\big[\mathbf{y}_n^{r\,\dagger}\mathbf{h}_n^r\big]\Big\} \\ &= \mathrm{Tr}\Big\{\mathbf{I}_M - \frac{\rho}{M}\bar{\mathbf{X}}\big(\boldsymbol{\Gamma} + \frac{\rho}{M}\bar{\mathbf{X}}^\dagger\bar{\mathbf{X}}\big)^{-1}\bar{\mathbf{X}}^\dagger\Big\} \\ &= \mathrm{Tr}\Big\{\big(\mathbf{I}_M + \frac{\rho}{M}\bar{\mathbf{X}}\boldsymbol{\Gamma}^{-1}\bar{\mathbf{X}}^\dagger\big)^{-1}\Big\}, \end{aligned} \tag{5.8}$$

where (5.8) is obtained by applying the Woodbury matrix identity [1]. It follows that $\tilde{\mathbf{w}}$ is approximately white with a variance of

$$\sigma_{\tilde{\mathbf{w}}}^2 = 1 + \frac{\rho \cdot \|\mathbf{x}\|}{M}\mathrm{Tr}\Big\{\big(\mathbf{I}_M + \frac{\rho}{M}\bar{\mathbf{X}}\boldsymbol{\Gamma}^{-1}\bar{\mathbf{X}}^\dagger\big)^{-1}\Big\}. \tag{5.9}$$

Next, based on (5.6) and (5.9), treating $\hat{\mathbf{H}}$ as the channel matrix, and $\sigma_{\tilde{\mathbf{w}}}^2$ as the effective noise, we can follow the same principle of the MCMC detector described in Chapter 3 and 4 to perform MIMO detection for each column of $\mathbf{X}$. The density function used for the GS to draw samples can be expressed as

$$p(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}}) = \frac{1}{(\pi\sigma_{\tilde{\mathbf{w}}}^2)^N}\exp\left\{-\frac{1}{\sigma_{\tilde{\mathbf{w}}}^2}\left\|\mathbf{y} - \sqrt{\frac{\rho}{M}}\hat{\mathbf{H}}\mathbf{x}\right\|\right\}. \tag{5.10}$$

### 5.4.2 Decorrelation SCE-MCMC Detector (DEC-MCMC)

In this section, we propose an improved version of SCE-MCMC to address the issue of error propagation arising from joint data detection and channel estimation. Recall

[1] $(\mathbf{A}+\mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{U}\mathbf{A}^{-1}$

that in SCE-MCMC, the APPs of all the data symbols in $\mathbf{X}$ are used to compute $\hat{\mathbf{H}}$. When the APPs are erroneous, the resulting $\hat{\mathbf{H}}$ is biased towards these wrong data decisions. Subsequently, when performing MCMC detection, the GS generates random samples according to a conditional density function that is determined by $\hat{\mathbf{H}}$ and the same set of APPs. The fact that these erroneous APPs are used twice in SCE-MCMC causes error propagation and thus performance degradation.

This motivates the design of a new detector, termed the decorrelation SCE-MCMC detector (DEC-MCMC), designed to reduce the correlation between the channel estimation and the data samples generated by the GS. It operates as follows. Consider an arbitrary column of $\mathbf{X}$, say $\mathbf{x}_t$ for $t \in 1, \cdots, T$. Let $\mathbf{X}_{-t}$ be the signal matrix obtained by removing column $\mathbf{x}_t$ from $\mathbf{X}$. When detecting symbols in $\mathbf{x}_t$, the channel estimate used by the MCMC detector is computed using the APPs of symbols in $\mathbf{X}_{-t}$. Specifically, in (5.4), (5.5), and (5.9), we replace $\mathbf{X}$, $\bar{\mathbf{X}}$, and $\mathbf{\Gamma}$ by $\mathbf{X}_{-t}$, $\bar{\mathbf{X}}_{-t}$, and $\mathbf{\Gamma}_{-t}$, respectively, to obtain the channel estimate and the variance of the effective noise as follows:

$$\hat{\mathbf{h}}_n^r = \sqrt{\frac{\rho}{M}}\mathbf{y}_n^r \mathbf{\Gamma}_{-t}^{-1}\bar{\mathbf{X}}_{-t}^{\dagger}\big(\mathbf{I}_M + \frac{\rho}{M}\bar{\mathbf{X}}_{-t}\mathbf{\Gamma}_{-t}^{-1}\bar{\mathbf{X}}_{-t}^{\dagger}\big)^{-1}, \tag{5.11a}$$

$$\sigma_{\tilde{\mathbf{w}}}^2 = 1 + \frac{\rho}{M}\mathrm{Tr}\big\{\big(\mathbf{I}_M + \frac{\rho}{M}\bar{\mathbf{X}}_{-t}\mathbf{\Gamma}_{-t}^{-1}\bar{\mathbf{X}}_{-t}^{\dagger}\big)^{-1}\big\}. \tag{5.11b}$$

The DEC-MCMC is presented in Algorithm 5.1. By DEC-MCMC, we ensure that the APPs of symbols in $\mathbf{x}_t$ do not contribute to the channel estimate that will be used to detect these symbols, which effectively alleviates the problem of error propagation. In Section 5.6, we provide numerical results to compare the performance of SCE-MCMC and DEC-MCMC. It is shown that a key factor that contributes to the superior performance of DEC-MCMC to SCE-MCMC is that the estimate of the effective noise by DEC-MCMC, i.e., (5.11b), is more accurate than that of SCE-MCMC, i.e., (5.9). The latter tends to under-estimate the effective noise because of the high correlation between the samples generated by the GS and the channel estimate. We also note that the advantage of DEC-MCMC over SCE-MCMC is more pronounced for small or moderate coherence lengths, in which case the contribution of the APPs of a single column in $\mathbf{X}$ to the channel estimate is more significant. These two detectors yield similar performance for larger coherence lengths due to the weaker correlation between channel estimation and the random sample generation.

**Algorithm 5.1:** DEC-MCMC algorithm

**Input**: $n$, $M$, $M_c$, $\mathbf{y}$, $\boldsymbol{\mu}_2^e$, $\mathcal{A} = \{a_0, \cdots, a_{2^{M_c}-1}\}$;

**begin**

Initialization: Obtain $\bar{\mathbf{X}}$ according to the APP as in (1.26);

**for** $t = T_p + 1$ *to* $T$ **do**

Set $\bar{x}_{m,t} = 0$ for all $m \in \{1, \cdots, M\}$ tentatively ;

Estimate $\hat{\mathbf{H}}_t$ and $\sigma^2_{\tilde{\mathbf{w}}}$ by (5.11a ) and (5.11b);

Run $G$ parallel GSs (as shown in Algorithm 3.3) for $\mathbf{x}_t$ to obtain a subset $\mathcal{L}_t$;

Calculate APP of $x_{m,t}$ for all $m \in \{1, \cdots, M\}$ and $j \in \{0, 2^{M_c} - 1\}$ by

$$\mu_1(x_{m,t} = a_j) = P(x_{m,t} = a|\mathbf{y}_t, \hat{\mathbf{H}}_t, \boldsymbol{\mu}_2^e) \propto \sum_{\substack{\mathbf{x}_t \in \mathcal{L}_t \\ x_{m,t} = a_j}} \exp\big\{ - \|\mathbf{y}_t - \sqrt{\frac{\rho}{M}}\hat{\mathbf{H}}_t\mathbf{x}_t\|\big\} P(x_{m,t} = a_j|\boldsymbol{\mu}_2^e); \tag{5.12}$$

**end**

**end**

### 5.4.3 Adaptive DEC-MCMC (ADA-MCMC) Detector

In this section, we propose an adaptive version of the DEC-MCMC detector to further reduce the complexity of MCMC detection. The complexity of the MCMC detector is determined by the total number of random samples generated by the GS and the choices of $G$ and $I$ are closely related to the modulation size and the channel estimation error. In general, under challenging detection scenarios, e.g., when the modulation size is large and the channel estimation is inaccurate, we should choose larger values of $G$ and $I$, thus collecting more random samples, in order to achieve satisfactory detection performance. For large modulation sizes, such as 64QAM, we have observed that it is necessary to choose $G = I = 20$ at the beginning iteration, when the channel estimation is most inaccurate, to optimize the performance of DEC-MCMC, whereas $G = I = 10$ is sufficient for 16QAM for the same detector. To reduce the complexity further for higher order modulations, we propose an adaptive DEC-MCMC detector, termed the ADA-MCMC.

The main idea of ADA-MCMC is to adaptively choose $G$ and $I$ based on the mean-square-error (MSE) of the channel estimate, i.e., select larger values of $G$ and

$I$ when the MSE is large, and reduce $G$ and $I$ as the MSE decreases over iterations. For simplicity, we set $G = I$ and describe how to adaptively choose $I(m)$, which is the $I$ value at the $m$-th iteration of joint data detection and channel decoding.

Let $\sigma^2(m)$ denote the MSE of the channel estimate at iteration $m$, and $I(m)$ be a linear function of $\sigma(m)$ such that $I(m) = k\sigma(m) + b$. The constants $k$ and $b$ are determined by considering the $I$ values for two special cases.

1. At the initial iteration, the MSE is the largest due to pilot-only channel estimation. The MSE is given by $\sigma_p^2 = M/(M + \rho T_p)$ [29, 69]. This case requires a maximum $I$ value, and we set it empirically to be $I_{\max}$.

2. In a genie-aided case, all the data decisions within a data block are error-free and are all used for channel estimation. It gives the smallest channel estimation error $\sigma_g^2 = M/(M + \rho T)$. This case requires a minimum $I$ value, and we set it to be $I_{\min}$.

We then substitute $I_{\max}$ and $I_{\min}$ into the linear function $I(m) = k\sigma(m) + b$ to yield $I_{\max} = k\sigma_p + b$ for the beginning iteration, and $I_{\min} = k\sigma_g + b$ for the genie-aided case. By solving these two equations, we obtain $k = \frac{I_{\max} - I\min}{\sigma_p - \sigma_g}$ and $b = I_{\max} - k\sigma_p$.

We evaluate the complexity saving of ADA-MCMC against DEC-MCMC using the number of samples generated by the GS. Compared to DEC-MCMC with fixed parameters $G_{\max}$ and $I_{\max}$, the percentage of the complexity saving of ADA-MCMC at iteration $m$ is given by $\eta(m) = \left(1 - \frac{G(m)I(m)}{G_{\max}I_{\max}}\right) \times 100\%$. By letting $\bar{\eta} = \frac{1}{L}\sum_{m=1}^{L} \eta(m)$, where $L$ is the total number of iterations, we obtain the average complexity saving $\bar{\eta}$. In Section 6.6, we demonstrate that significant complexity saving of $\bar{\eta}$ up to 60% can be achieved by ADA-MCMC with negligible performance loss.

## 5.5 Genie-aided Channel Estimation MCMC (GAD-MCMC) as a Performance Benchmark

While it is common to use the receiver performance under perfect CSI as a performance benchmark, for large MIMO systems with higher order modulations, it becomes difficult to approach this idealized benchmark due to the increased challenge in data detection and channel estimation. In this work, we introduce a genie-aided

channel estimation MCMC (GAD-MCMC) as a more realistic performance benchmark for the proposed design. It works as follows.

When detecting an arbitrary vector $\mathbf{x}_t$ in $\mathbf{X}$, we assume that all the symbols in $\mathbf{X}_{-t}$ are perfectly known. Accordingly, a LMMSE channel estimate can be computed based on $\mathbf{X}_{-t}$ and then used to detect $\mathbf{x}_t$. This represents the best channel estimation possible for $\mathbf{x}_t$ when all the remaining vectors in $\mathbf{X}$ are known[2]. Thus, GAD-MCMC provides a realistic performance bound for joint data detection, channel estimation, and data decoding.

GAD-MCMC can be considered as an idealized case of DEC-MCMC. In fact, the performance of DEC-MCMC should approach that of GAD-MCMC as the quality of the soft information provided by the channel decoder improves over iterations. This is confirmed in Section 6.6 in which we show that DEC-MCMC can indeed approach the performance of GAD-MCMC for a variety of channels considered in this work.

## 5.6 Simulation Results

In this section, we present simulation results to examine the performance of the proposed receiver design. Performance of SCE-MCMC, DEC-MCMC, and ADA-MCMC are provided for block fading channels with various coherence lengths. Comparisons of the proposed MCMC detectors with an improved version of the MMSE detector is presented to demonstrate the substantial performance gain of the MCMC detectors.

### 5.6.1 Comparisons of MCMC Detectors with Imperfect CSI

We first consider a $3 \times 3$ MIMO system with 16QAM modulation. A rate $R = 1/2$ convolutional code with generator polynomial $(117, 155)_{octal}$ is used. The code length is 8000 bits. This system has a spectral efficiency of 12 bits/channel use. In Fig. 5.4, we compare performance of DEC-MCMC, SCE-MCMC, together with two benchmark curves corresponding to GAD-MCMC and PSA-MCMC. The channel estimation in the latter employs the PSAM method, which performs only once at the first iteration,

[2]While it is possible to obtain a better channel estimate by assuming that the entire $\mathbf{X}$ is known, and then use it for data detection, this approach suffers from data over-fitting, which leads to a receiver performance that is even better than the case of perfect CSI.

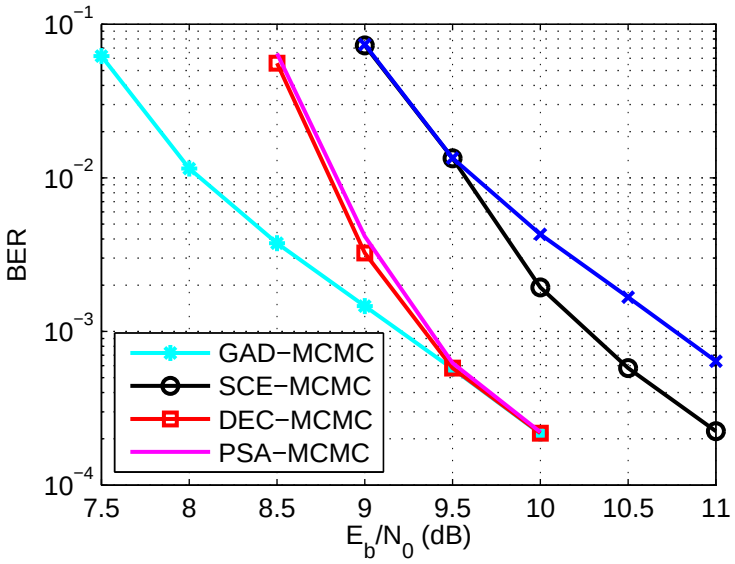


**Figure 5.4**. Performance comparison for a $3 \times 3$ MIMO system with 16QAM, $T = 10$, $G = I = 10$.

and remains the same in the successive iterations. The MCMC parameters are $G = I = 10$. A total of ten iterations of joint MCMC detection, channel estimation, and data decoding is performed at the receiver. As shown in Fig. 5.4, DEC-MCMC achieves the best performance. It performs closely to the GAD-MCMC at the bit-error rate (BER) of about $10^{-3}$. It outperforms SCE-MCMC and PSA-MCMC by about 1 dB and 1.5 dB, respectively. The 1 dB gain of DEC-MCMC over SCE-MCMC justifies the importance of reducing the correlation between channel estimation and data decision. We note that the gap between PSA-MCMC and DEC-MCMC will increase further as $T$ increases.

To better understand the performance gap between DEC-MCMC and SCE-MCMC, we examine the accuracy of the estimated effective noise variance, $\sigma^2_{\tilde{\mathbf{w}}_t}$, defined in (5.11b) for DEC-MCMC. The estimated effective noise variance for SCE-MCMC can be found by replacing $\mathbf{X}_{-t}$ and $\mathbf{\Gamma}_{-t}$ in (5.11b) by $\mathbf{X}$ and $\mathbf{\Gamma}$. This quantity directly affects the quality of random samples generated by the GS. In Fig. 5.5, at $E_b/N_0 = 9.5$ dB, we plot $\sigma^2_{\tilde{\mathbf{w}}}$, obtained by taking the average of $\sigma^2_{\tilde{\mathbf{w}}_t}$ over $M+1 \le t \le T$, and compare it with the actual effective noise variance $\sigma^2_{\tilde{\mathbf{w}},\text{act}} = 1 + \frac{\rho}{M}\|\mathbf{x}\| \sum_{1\le m\le M, 1\le n\le N} |\hat{h}_{n,m} - h_{n,m}|^2$, as a function of iterations. For SDD-MCMC, due to the correlation be-

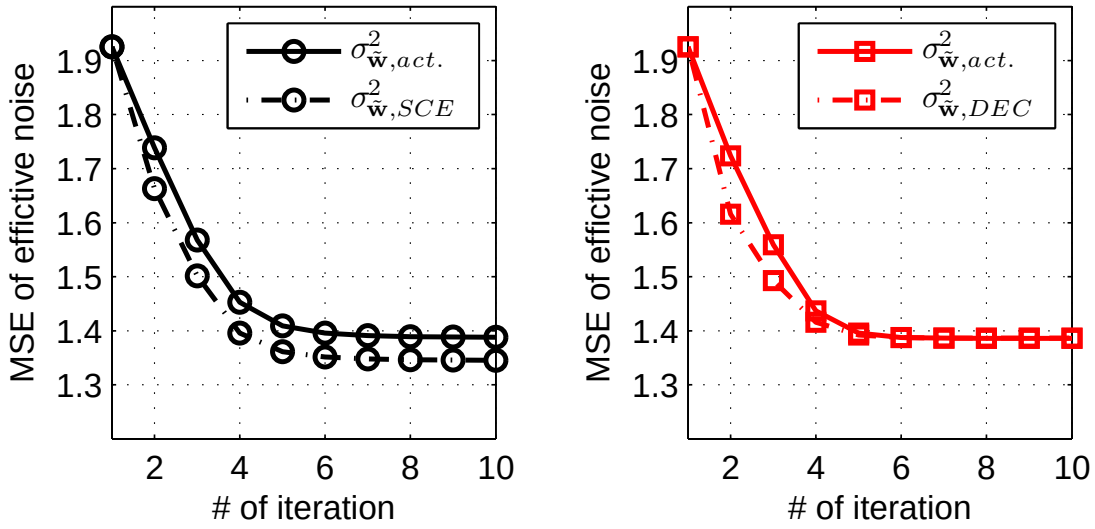


**Figure 5.5**. Difference between actual MSE and estimated MSE of effective noise for a $3 \times 3$ MIMO system with 16QAM at $E_b/N_0 = 9.5$ dB. $T = 10$.

tween data and channel estimation, $\sigma^2_{\tilde{\mathbf{w}}}$ fails to converge to $\sigma^2_{\tilde{\mathbf{w}},\text{act}}$ and remains an under-estimate. For DEC-MCMC, $\sigma^2_{\tilde{\mathbf{w}}}$ converges to $\sigma^2_{\tilde{\mathbf{w}},\text{act}}$ at later iterations due to decorrelation. This explains the superior performance of DEC-MCMC.

In Fig. 5.6, we examine the performance of the proposed schemes for a $4 \times 4$ MIMO system with 64QAM modulation. This system has a higher spectral efficiency

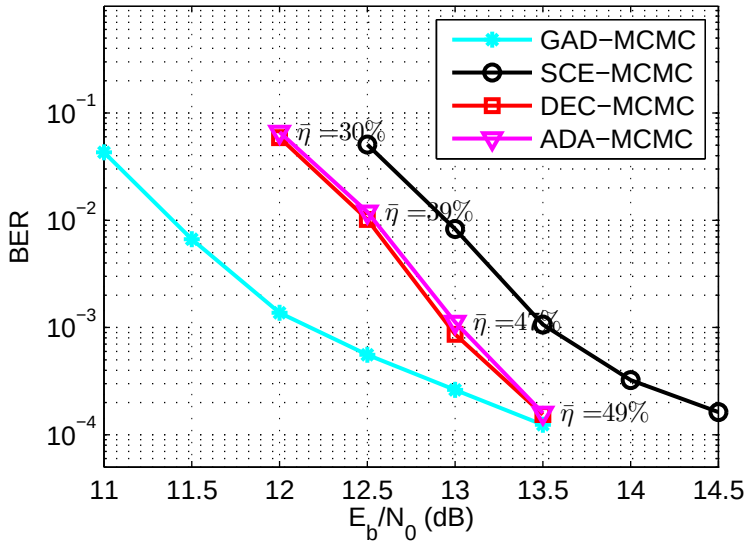


**Figure 5.6**. Performance comparison for a $4 \times 4$ MIMO system with 64QAM, $T = 10$. $G = I = 20$ for GAD/SCE/DEC-MCMC. $G_{\max} = I_{\max} = 20$ and $G_{\min} = I_{\min} = 10$ for ADA-MCMC.

of 24 bits/channel use. Due to the large system dimension, we increase the MCMC parameters to $G = I = 20$ and adopt the ADA-MCMC to further reduce the detection complexity. As shown in Fig. 5.6, the best performance is still achieved by DEC-MCMC, which approaches the performance of GAD-MCMC at the BER of $10^{-4}$. It is observed that DEC-MCMC outperforms SCE-MCMC by about 1.5 dB at BER$=10^{-4}$. For this system, PSA-MCMC outperforms SCE-MCMC by only $0.3 \sim 0.4$ dB, because the performance of SCE-MCMC is degraded by the strong correlation between the channel estimation and data decision. It is also shown in Fig. 5.6 that ADA-MCMC performs closely to DEC-MCMC at a reduced complexity.

A detailed complexity analysis is provided in Fig. 5.7. Here, for each of the four $E_b/N_0$ points on the ADA-MCMC curve in Fig. 5.6, we show that $\eta(m)$ increases as a function of $m$. This reveals that as the quality of channel estimation improves over iterations, the complexity saving becomes more significant. Note that $\eta(m)$ increases from 10% to about 60% after ten iterations. We also observe that $\eta(m)$ increases with $E_b/N_0$, due to superior channel estimation at higher SNR.

### 5.6.2 Performance Comparison with MMSE Detector

In this section, we compare performance of the proposed DEC-MCMC detectors with an improved version of the MMSE detector of [16]. We call this improved

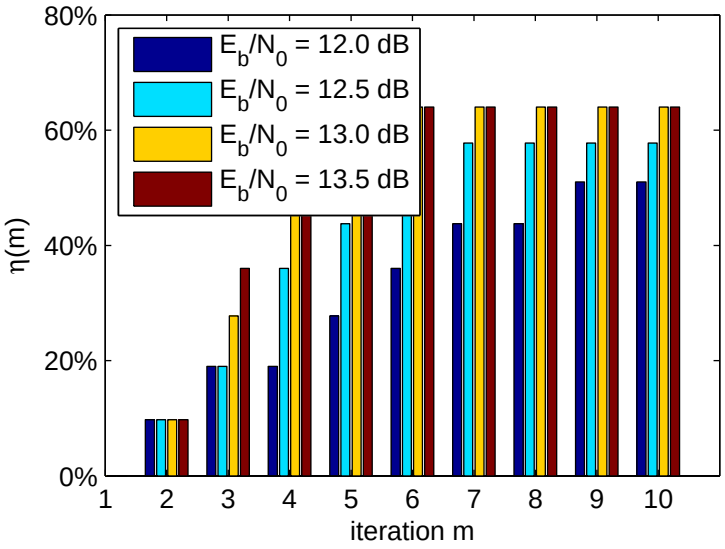


**Figure 5.7**. ADA-MCMC complexity save.

detector DEC-MMSE. We decorrelate the channel estimation and MIMO detection in the same way as that of DEC-MCMC. It turns out that this improved MMSE detector performs significantly better than its counterpart in [16].

In Fig. 5.8, we consider $4 \times 4$ MIMO systems with 64QAM modulation. Two larger $T$ values are considered: $T = 20$ and $T = 50$. Since the block lengths are longer than the previous cases, the BER curves are less likely to show an error floor at high SNR. It is shown that at BER $= 10^{-4}$, DEC-MCMC provides a substantial gain over DEC-MMSE by 1.5 dB for $T = 20$, and by about 1.7 dB for $T = 50$. The complexity saving of ADA-MCMC over DEC-MCMC remains significant, with $\bar{\eta}$ ranging from 42% to 62%, at the cost of a slight performance degradation. The gap between ADA-MCMC and DEC-MCMC increases to about 0.25 dB for $T = 50$ at BER $= 10^{-4}$. The performance of SCE-MCMC (not shown in the figure) is similar to that of DEC-MCMC for these two $T$ values because the contribution of the data decisions of a single column towards SDD-CE diminishes as $T$ increases.

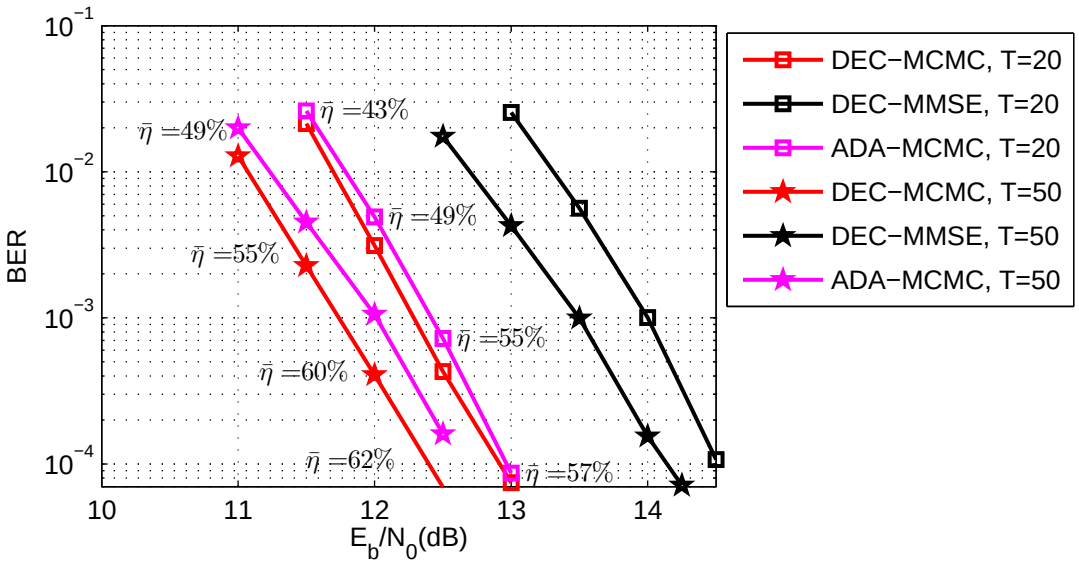


**Figure 5.8**. Performance comparison for a $4 \times 4$ MIMO system with 64QAM. $G = I = 20$ for GAD/SCE/DEC-MCMC. $G_{\max} = I_{\max} = 20$ and $G_{\min} = I_{\min} = 10$ for ADA-MCMC.

## 5.7 Conclusion

In this chapter, we developed joint MCMC detection and SCE algorithms that achieved excellent performance for MIMO systems with high spectral efficiency. The proposed algorithms are applicable to general communication scenarios with large number of transmit antennas, and arbitrary modulation size. An analytical study of the SCE algorithm was presented, based on which, new MCMC detectors that explicitly take the channel estimation error into account were designed to facilitate joint iterative data detection and channel estimation. The proposed DEC-MCMC effectively reduces the correlation between channel estimation and data detection, and thus yields a substantial performance gain for moderate or fast fading scenarios. The ADA-MCMC is a novel adaptive MCMC detector that controls the complexity of MCMC detection according to the channel estimation error. The proposed design demonstrated significant performance gain over the soft MMSE detector for a variety of channels considered in this work.

# CHAPTER 6

# JOINT CHANNEL ESTIMATION AND MCMC DETECTOR IN TIME-SELECTIVE RAYLEIGH FADING CHANNELS

In Chapter 5, we considered a channel model that remains constant over each block of data and changes to another, independent, channel for the next data block. This, of course, is unrealistic and is merely of interest to see how a given receiver performs close to the theoretical capacity bound. In this chapter, we consider a more realistic channel model that continuously varies with time. We also introduce novel ideas that using the estimated fading statistics introduces negligible performance loss compared to knowing the fading statistics exactly.

## 6.1 Introduction

Multiple-input multiple-out (MIMO) technique improves the wireless link reliability and/or increases the channel capacity by a factor equal to the minimum number of the transmit/receive antennas [2]. In the past, a vast amount of research has been performed to address MIMO receiver design related to high data-rate and high mobility in wireless communication systems [3, 11, 41, 45, 46, 52, 70]. In the absence of channel state information (CSI), the receiver needs to provide a satisfying channel estimate for coherent MIMO detection. For instance, conventional pilot symbol assisted modulation (PSAM) channel estimation [29,31,32] is widely used and discussed because of its simplicity for implementation. However, the performance of PSAM channel estimation is limited by the power and number of channel uses spent on training symbols.

More powerful channel estimators can be developed by taking advantage of the so-called *turbo principle* and devising methods that detect data and channel iteratively.

Such systems may generically be referred to as *turbo receivers*. Turbo receivers refine the estimated channel and the quality of the detected data symbols as the iterations proceed. Early work on turbo receivers is due to Valenti and Woerner [35, 71], where their study is performed for the case where data symbols are from binary phase-shift keying (BPSK) modulation and the channel is single-input single-output and frequency-flat fading. An extension of the turbo receiver of [35, 71], for the case where data symbols are from a QAM constellation, is presented in [36]. It has been observed in [36] that this results in a very high complexity, as the underlying correlation matrix of the channel estimator is data dependent and thus its inverse has to be calculated at every channel use.

The goal of this chapter is to develop a low-complexity channel estimation algorithm for turbo receivers, aiming to approximate the near-optimal performance achieved by Wiener filtering at a much reduced complexity. We consider time-varying, frequency flat channels and examine the performance of the proposed channel estimation algorithm for MIMO communications with higher order modulations. We propose a *dual-layer* Wiener filtering (DLWF) channel estimation algorithm. In the first layer, the proposed algorithm computes a sequence of coarse channel estimates based on a block-fading approximation. Subsequently, a smoothing filter is used in the second layer to refine these coarse channel estimates. The smoothing filter is a time-invariant Wiener filter whose coefficients are derived based on the second-order statistics of the channel estimates.

We examine the performance of the proposed dual-layer channel estimation algorithm using two state-of-the-art MIMO detectors: the soft-MMSE detector and the Markov chain Monte Carlo (MCMC) detector. For large MIMO systems with a high spectral efficiency of at least 10 bits/channel use, our results demonstrate that the proposed DLWF algorithm, in conjunction with the MCMC-MIMO detector, achieves a performance very close to that of the optimal WF filtering with a significant complexity reduction. The MMSE detector, on the other hand, suffers a substantial performance loss compared to that of the MCMC detector, when the same channel estimation algorithm is applied. In this work, we also propose a simple method to estimate the fading rate of time-varying channels. Computer simulation results

reveal that the proposed fading rate estimator incurs only a negligible performance loss compared to the case where the fading rate is perfectly known.

We note that similar ideas of dual-layer channel estimation exist in the literature [35,36,38,39]. These work use short blocks of data decisions to obtain coarse estimates of the channel at each channel use, and then these channel estimates are smoothed through a Wiener filter for further refinement. The key differences between the proposed DLWF and those of [35, 36, 38, 39] are as follows. First, the coarse channel estimates obtained in [35, 36, 38, 39] are based on hard decisions. This suffers from the problem of error propagation. In this work, we propose to use soft estimates of the data symbols. Thus, our algorithm is able to compensate for decision errors as well as additional errors caused by channel variation. Second, in [35, 36, 38, 39], the filter coefficients used in the second layer ignore the coarse channel estimation error. In this work, the smoothing filter designed in the second layer takes into account the channel estimation error. Our simulation studies, presented in Section 6.6, show that the proposed design yields a second layer filter that is better matched to the first layer channel estimates; hence, a significant improvement in the receiver performance is observed.

The rest of this chapter is organized as follows. In Section 6.2, we introduce the channel model. The optimal (linear/Wiener filter) channel estimator is presented in Section 6.3. The proposed low-complexity dual-layer channel estimation algorithm is presented in Section 6.4. In Section 6.5, we present a simple method to estimate the fading rate of the time-varying channel. Simulation results are presented in Section 6.6. The conclusions of the chapter are drawn in Section 6.7.

## 6.2 System Setup

We consider a frequency-flat time-varying MIMO channel with $M$ transmit antennas and $N$ receive antennas. A block diagram of the proposed system is shown in Fig. 6.1.

An information bit sequence $\mathbf{b}$ is encoded by a channel encoder of rate $R$. The coded bit sequence $\mathbf{d}$ is interleaved according to a permutation function $\Pi(\cdot)$ and then mapped to a symbol sequence using a $2^{M_c}$-ary constellation. After adding $T_p$

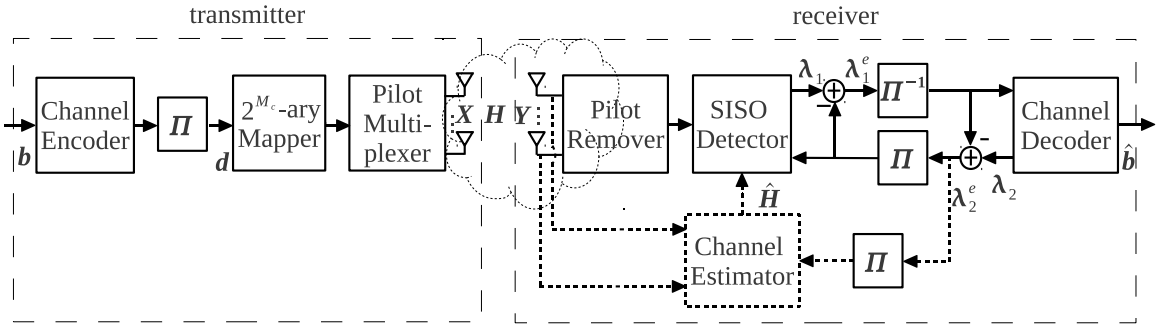


**Figure 6.1**. The block diagram of proposed MIMO system.

pilot symbols in every $T_d = T_c - T_p$ data symbols on each antenna, the resulting symbol sequences, denoted by $\mathbf{x}_{m,t}$ for $m \in [1, M]$ and $t \in [1, L_{\text{packet}}]$, are launched by $M$ transmit antennas. Here, the minimum training sequence $T_p = M$ is used to maximize the bandwidth efficiency. An average power constraint is imposed such that $\frac{1}{M}\mathrm{E}[\mathbf{x}_t^\dagger \mathbf{x}_t] = 1$, where $\mathbf{x}_t \in \mathcal{C}^{M\times 1}$ denotes the symbols transmitted at time $t$.

The signal received by the $n$-th antenna at time instant $t$ is

$$y_{n,t} = \sqrt{\frac{\rho}{M}}\mathbf{h}_{n,t}^{\mathrm{T}}\mathbf{x}_t + w_{n,t}, \quad \text{for } 1 \le n \le N \tag{6.1}$$

where $\mathbf{h}_{n,t} \in \mathcal{C}^{M\times 1}$ denotes the channel gains between all transmit antennas and the $n$-th receive antenna, $\rho$ is the signal-to-ratio (SNR) at each receive antenna, and $w_{n,t}$ is a white complex Gaussian noise with zero mean and unit variance. We also define the MIMO channel gain matrix $\mathbf{H}_t = \left[\mathbf{h}_{1,t}\ \mathbf{h}_{2,t}\ \cdots\ \mathbf{h}_{N,t}\right]$. It is assumed that the elements of $\mathbf{H}_t$, $h_{mn,t}$ for $m \in [1, M]$ and $n \in [1, N]$, are a set of independent and identically distributed (i.i.d.) complex Gaussian variables with zero mean and unit variance, and have the correlation coefficients $\gamma_\tau = \mathrm{E}[h_{mn,t}h_{mn,t+\tau}^*]$ for $\tau = 0, \pm 1, \pm 2, \cdots$. Due to the absence of spatial correlation, we drop the antenna index $n$ in the sequel for simplicity.

At the receiver, MIMO detection, channel estimation, and data decoding are performed in an iterative manner, i.e., in a turbo loop. We let $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ denote the *a posteriori probabilities* (APPs) of the transmitted symbols produced by the MIMO detector, and by the channel decoder, respectively. The corresponding extrinsic

information is denoted by $\boldsymbol{\lambda}^e$ and $\boldsymbol{\mu}^e$, respectively. At each iteration, given $\boldsymbol{\mu}$, the mean and variance of each transmitted symbol $x_{m,t}$, denoted by $\bar{x}_{m,t}$ and $v_{m,t}$, respectively, are obtained as

$$\bar{x}_{m,t} = \sum_{a \in \mathcal{A}} a \cdot P(x = a|\boldsymbol{\mu}), \quad v_{m,t} = \sum_{a \in \mathcal{A}} |a|^2 \cdot P(x = a|\boldsymbol{\mu}) - |\bar{x}|^2 \tag{6.2}$$

where $\mathcal{A}$ denotes the set of data symbol constellation points. Then, channel estimation is performed to obtain $\hat{\mathbf{H}}_t$ according to the statistics of soft decisions and the received signal sequences. Subsequently, $\hat{\mathbf{H}}_t$ and $\boldsymbol{\mu}^e$ are fed to the MIMO detector for data detection. The MIMO detector generates updated APPs $\boldsymbol{\lambda}$, and then $\boldsymbol{\lambda}^e$ is passed to the channel decoder. In this way, joint MIMO detection, channel estimation, and data decoding is performed iteratively. After a fixed number of iterations, decisions are made to obtain the estimated bit sequence $\hat{\mathbf{b}}$.

## 6.3 Optimal Channel Estimator

In prior work [35, 38], a Wiener filter channel estimator was derived for turbo receivers by taking the decision values at the channel decoder output as the actual transmit symbols, i.e., ignoring possible errors in the decisions. This leads to some performance degradation compared to Wiener filter channel estimators that use soft decisions [33, 40]. While [33, 40] considered only MIMO systems with BPSK modulation, we can follow a similar approach to derive a Wiener filter channel estimator for higher order modulations. For the completeness of the chapter, this derivation is presented in this section.

We seek to find an estimate of the channel gain $\mathbf{h}_t$ given the soft values (equivalently, APPs) of the transmitted symbols and the correlation coefficients $\gamma_\tau$. In the sequel, we drop the antenna index $n$, and hence, consider the channel equation

$$y_t = \sqrt{\frac{\rho}{M}} \mathbf{h}_t^{\mathrm{T}} \mathbf{x}_t + w_t = \sqrt{\frac{\rho}{M}} \mathbf{x}_t^{\mathrm{T}} \mathbf{h}_t + w_t. \tag{6.3}$$

To develop an estimator for $\mathbf{h}_t$, the closet $2K + 1$ samples of the received signal surrounding time $t$ are considered here. Accordingly, we define the received signal vector $\mathbf{y}_t^{(K)} = [y_{t-K} \ \cdots y_t \ \cdots \ y_{t+K}]^{\mathrm{T}}$ and note that

$$\mathbf{y}_t^{(K)} = \sqrt{\frac{\rho}{M}} \mathbf{S}_t \mathbf{g}_t + \mathbf{w}_t, \tag{6.4}$$

where $\mathbf{g}_t = [\mathbf{h}_{t-K}^{\mathrm{T}} \ \cdots \ \mathbf{h}_t^{\mathrm{T}} \ \cdots \ \mathbf{h}_{t+K}^{\mathrm{T}}]^{\mathrm{T}}$ is an $M(2K+1) \times 1$ column vector obtained by stacking channel vectors $\mathbf{h}_k$, for $t-K \leq k \leq t+K$, in a column, and $\mathbf{S}_t$ is a $(2K+1) \times M(2K+1)$ matrix of the transmitted symbols defined as

$$\mathbf{S} = \begin{bmatrix} \mathbf{x}_{t-K}^{\mathrm{T}} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{x}_t^{\mathrm{T}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{x}_{t+K}^{\mathrm{T}} \end{bmatrix}. \tag{6.5}$$

Here, $\mathbf{0}$ is the $1 \times M$ zero vector.

The optimal channel estimator is a linear MMSE (LMMSE) filter [72] with the input $\mathbf{y}_t$ and coefficient matrix $\mathbf{F}_t = \begin{bmatrix} \mathbf{f}_{t,1} & \mathbf{f}_{t,2} & \cdots & \mathbf{f}_{t,M} \end{bmatrix}$. The estimated channel vector is that obtained as

$$\hat{\mathbf{h}}_t = \mathbf{F}_{\mathrm{o},t}^{\dagger} \mathbf{y}_t^{(K)}, \tag{6.6}$$

where the superscript $^{\dagger}$ denotes Hermitian and the subscript 'o' is added to emphasize that $\mathbf{F}_{\mathrm{o},t}$ is the optimum choice of $\mathbf{F}_t$. Following standard derivations of the LMMSE estimator [33], it is straightforward to show that

$$\mathbf{F}_{\mathrm{o,t}} = \mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}^{-1} \mathbf{\Theta}_{\mathbf{y}_t\mathbf{h}_t}. \tag{6.7}$$

where $\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}^{-1} = \mathrm{E}[\mathbf{y}_t^{(K)}(\mathbf{y}_t^{(K)})^{\dagger}]$ and $\mathbf{\Theta}_{\mathbf{y}_t\mathbf{h}_t} = \mathrm{E}[\mathbf{y}_t^{(K)}\mathbf{h}_t^{\dagger}]$, and $\mathrm{E}[\cdot]$ denotes expectation. An important point to note here is that the matrices $\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}$ and $\mathbf{\Theta}_{\mathbf{y}_t\mathbf{h}_t}$ are data dependent and thus vary with time. Hence, $\mathbf{F}_{\mathrm{o,t}}$ must be calculated/updated at each time instant. For large values of $K$, this can result in a very high complexity, particularly if one notes that the $(2K+1) \times (2K+1)$ matrix $\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}$ must be inverted for every $t$.

For completeness of our derivation here, we note that

$$\mathbf{\Theta}_{\mathbf{y}_t\mathbf{h}_t} = \sqrt{\frac{\rho}{M}} \mathbf{\Upsilon}_t^{(K)} \mathbf{X}_t^{(K)}, \tag{6.8}$$

where $\mathbf{\Upsilon}_t^{(K)}$ is the $(2K+1) \times (2K+1)$ diagonal matrix with the elements of $\gamma_{-K}$, $\cdots$, $\gamma_0$, $\cdots$, $\gamma_K$, and $\mathbf{X}_t^{(K)} = [\mathbf{x}_{t-K} \ \cdots \ \mathbf{x}_t \ \cdots \ \mathbf{x}_{t+K}]^{\mathrm{T}}$. Also, the elements of $\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}$ are given by

$$\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}(i,j) = \frac{\rho}{M} \gamma_{i-j} \cdot \langle \mathbf{x}_{i+t-K}, \mathbf{x}_{j+t-K} \rangle + \delta(i-j), \tag{6.9}$$

where $\langle \mathbf{x}_{i+t-K}, \mathbf{x}_{j+t-K} \rangle$, for $i, j \in [0, 2K]$, denotes the inner product of $\mathbf{x}_{i+t-K}$ and $\mathbf{x}_{j+t-K}$, and $\delta(t)$ is the Kronecker-delta function.

The above LMMSE estimator assumes that the transmit data symbols $\mathbf{x}_t$ are known. This, of course, cannot be true as the main goal of the receiver is obtain estimates of the transmitted data. The works presented in [35, 38] take the decision values at the channel decoder output as the actual transmitted symbols, i.e., ignoring possible errors in the decisions. This inevitably leads to some loss in performance of the receiver. A more sophisticated approach is to use the soft decisions from the channel decoder instead of $\mathbf{x}_t$ and modify the LMMSE estimator accordingly. As discussed previously, this approach has been taken by some authors, e.g., [33, 40], where it is noted that the LMMSE estimator gain matrix is given by

$$\bar{\mathbf{F}}_{\mathrm{o,t}} = \bar{\mathbf{\Phi}}^{-1}_{\mathbf{y}_t\mathbf{y}_t} \bar{\mathbf{\Theta}}_{\mathbf{y}_t\mathbf{h}_t}, \tag{6.10}$$

where $\bar{\mathbf{\Theta}}_{\mathbf{y}_t h_t}$ has the same form as $\mathbf{\Theta}_{\mathbf{y}_t\mathbf{h}_t}$ in (6.8) with $\mathbf{x}_{t-K}$ through $\mathbf{x}_{t+K}$ replaced by $\bar{\mathbf{x}}_{t-K}$ through $\bar{\mathbf{x}}_{t+K}$, respectively, calculated according to (6.2). Also, the elements of $\bar{\mathbf{\Phi}}_{\mathbf{y}_t,\mathbf{y}_t}$ are found as

$$\bar{\mathbf{\Phi}}_{\mathbf{y}_t\mathbf{y}_t}(i,j) = \frac{\rho}{M}\gamma_{i-j}\langle \bar{\mathbf{x}}_{i+t-K}, \bar{\mathbf{x}}_{j+t-K}\rangle + (1 + \frac{\rho}{M}\sum_{m=1}^{M} \upsilon_{m,i+t-K}) \cdot \delta(i-j). \tag{6.11}$$

In the rest of this chapter, we refer to (6.10), with the correlation matrix $\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}$ in (6.11), as the optimum Wiener filter (OWF) channel estimator. An important point to note here is that the matrices $\bar{\mathbf{\Phi}}_{\mathbf{y}_t\mathbf{y}_t}$ and $\bar{\mathbf{\Theta}}_{\mathbf{y}_t\mathbf{h}_t}$ are data dependent and thus vary with time. Hence, $\bar{\mathbf{F}}_{\mathrm{o,t}}$ must be calculated/updated at each time instant. For large values of $K$, this can result in a very high complexity, particularly if one notes that the $(2K+1) \times (2K+1)$ matrix $\mathbf{\Phi}_{\mathbf{y}_t\mathbf{y}_t}$ must be inverted for every $t$. On the other hand, to approach the best performance that one can obtain from the turbo receiver, a choice of $2K+1$ (the size of the above matrices) in the order of 100 is not unusual. This clearly implies that the implementation of the OWF channel estimator can be computationally very involved. Hence, development of near-optimum channel estimators (respectively, turbo receivers), but with a reasonable complexity, is of great interest. Such an algorithm is developed in the next section.

## 6.4 Dual-layer Channel Estimator

In this section, we describe the proposed dual-layer channel estimator. The key components of the proposed design are shown in Fig. 6.2. In the first layer, LMMSE

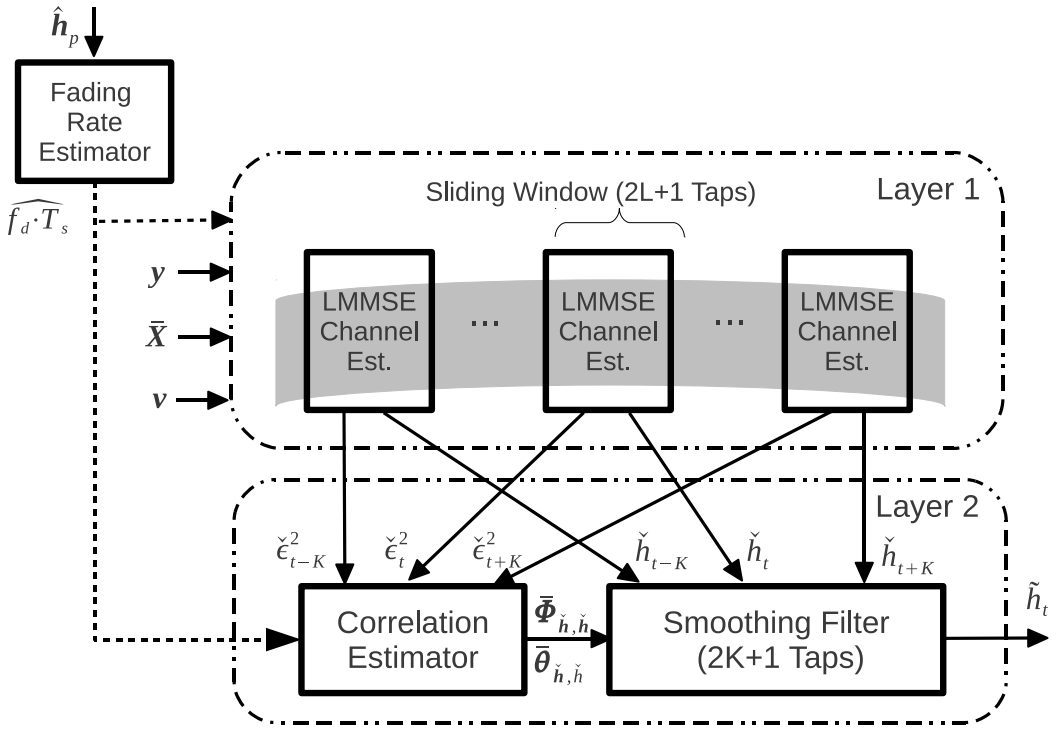


**Figure 6.2**. Block diagram of the dual-layer channel estimator.

channel estimation is performed over a sliding window of length $2L + 1$ to generate a coarse channel estimate for the channel use at the center of the sliding window. Subsequently, the sequence of coarse channel estimates generated from the first layer is passed to a smoothing filter in the second layer to refine the channel estimation. A distinct feature of our design lies in the correlation estimator in which we estimate the second-order statistics of the coarse channel estimates in order to derive the coefficients of the smoothing filter. This is in contrast to prior approaches [35, 36, 38, 39] in which such correlation is ignored. Furthermore, we note that the proposed design does not require prior knowledge of the fading rate. Satisfactory performance can be achieved using a simple fading rate estimator that we derive in Section 6.5. In the remainder of this section, we describe each of these key components in detail.

### 6.4.1 Layer 1: Coarse Channel Estimation

At this stage, for each time instant $t$, an estimate $\check{\mathbf{h}}_t$ of the channel gain $\mathbf{h}_t$ is obtained using the received signal samples $\mathbf{y}_t^{(L)} = [y_{t-L}, \cdots, y_t, \cdots, y_{t+L}]^{\mathrm{T}}$ and the system model

$$y_{t+l} = \sqrt{\frac{\rho}{M}}\mathbf{x}_{t+l}^{\mathrm{T}}\mathbf{h}_t + v_{t+l} + w_{t+l}, \quad \text{for } -L \le l \le L. \tag{6.12}$$

In (6.12), $v_{t+l} = \sqrt{\frac{\rho}{M}}\mathbf{x}_{t+l}^{\mathrm{T}}(\mathbf{h}_{t+l} - \mathbf{h}_t)$ is treated as an additional noise term (beside the channel additive white Gaussian noise $w_t$), and $L$ is an integer which is typically much smaller than the parameter $K$ of the OWF channel estimator. Beside the received signal samples, the information that we use to obtain $\check{\mathbf{h}}_t$ are (i) the soft symbol information from the channel decoder, summarized in (6.2); (ii) the channel statistics, characterized by the correlation coefficients $\gamma_\tau$; and (iii) the statistical characteristics of the additive noise $w_t$. Also, to simplify our derivations, we define $w'_t = v_t + w_t$ and, thus, rewrite (6.12) as

$$y_{t+l} = \sqrt{\frac{\rho}{M}}\mathbf{x}_{t+l}^{\mathrm{T}}\mathbf{h}_t + w'_{t+l}, \quad \text{for } -L \le l \le L. \tag{6.13}$$

Using (6.13), the estimator coefficient matrix, here, is obtained as

$$\check{\mathbf{F}}_t = \check{\mathbf{\Phi}}_t^{-1}\check{\mathbf{\Theta}}_t \tag{6.14}$$

where

$$\check{\mathbf{\Theta}}_t \triangleq \check{\mathbf{\Theta}}_{\mathbf{y}_t\mathbf{h}_t} = \sqrt{\frac{\rho}{M}}\bar{\mathbf{X}}_t^{(L)} \tag{6.15}$$

$$\check{\mathbf{\Phi}}_t \triangleq \check{\mathbf{\Phi}}_{\mathbf{y}_t\mathbf{y}_t} = \frac{\rho}{M}\bar{\mathbf{X}}_t^{(L)}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger} + \mathbf{Q}_t, \tag{6.16}$$

and $\mathbf{Q}_t$ is the correlation matrix of the additive noise $w'_{t+l}$ of (6.13) given by

$$\begin{aligned}\mathbf{Q}_t &= \mathbf{I}_{2L+1} + \frac{\rho}{M}\mathrm{E}\Big\{\big(\mathbf{X}_t^{(L)} - \bar{\mathbf{X}}_t^{(L)}\big)\big(\mathbf{X}_t^{(L)} - \bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\Big\} + \mathbf{\Psi}_t \\ &\approx \mathbf{I}_{2L+1} + \frac{\rho}{M}\mathrm{Diag}\{\sum_{m=1}^{M} v_{m,t-L}, \cdots, \sum_{m=1}^{M} v_{m,t+L}\} + \mathbf{\Psi}_t. \end{aligned} \tag{6.17}$$

Here, $\mathbf{X}_t^{(L)} = [\mathbf{x}_{t-L} \ \cdots \ \mathbf{x}_t \ \cdots \ \mathbf{x}_{t+L}]^{\mathrm{T}}$ consists of $2L+1$ rows of transmit data symbols, and $\bar{\mathbf{X}}_t^{(L)}$ is defined on the soft decision of $\mathbf{X}_t^{(L)}$ correspondingly. $\mathrm{Diag}\{a_1, a_2, \cdots\}$ represents a diagonal matrix with $a_1, a_2, \cdots$ on its main diagonal. In (6.17), $\mathbf{I}_{2L+1}$ is

the identity matrix of size $2L+1$, representing the correlation matrix of noise samples $w_t$; the second term arises from the decisions uncertainty, and it can be approximated as a diagonal matrix with $\frac{\rho}{M}\sum_{m=1}^{M} v_{m,i}$ for $i \in [t-L, t+L]$ on its main diagonal; and the third term, $\mathbf{\Psi}_t$, is the correlation matrix of the noise term $v_{t+l}$ of (6.12). In [73], for a Jakes' channel model, it has been shown that $\mathbf{\Psi}_t$ is a diagonal matrix with the $l$th diagonal element

$$\psi_l = \frac{\rho \mathrm{E}[\|\mathbf{x}_l\|^2]}{M} 2[1 - \mathrm{sinc}(2\pi f_d T_s l)]. \tag{6.18}$$

Summarizing the above results, we get

$$\check{\mathbf{h}}_t \;=\; \check{\mathbf{F}}_t^{\dagger}\mathbf{y}_t^{(L)} = \sqrt{\frac{\rho}{M}}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\left(\frac{\rho}{M}\bar{\mathbf{X}}_t^{(L)}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger} + \mathbf{Q}_t\right)^{-1}\mathbf{y}_t^{(L)} \tag{6.19}$$

where $\mathbf{y}_t^{(L)} = [y_{t-L} \;\cdots\; y_t \;\cdots\; y_{t+L}]^{\mathrm{T}}$. Computation of $\check{\mathbf{h}}_t$ involves inversion of the $(2L+1)\times(2L+1)$ matrix $\frac{\rho}{M}\bar{\mathbf{X}}_t^{(L)}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger} + \mathbf{Q}_t$ with the typical choice of $L = 10$ to 20. The complexity of implementation of (6.19) can be reduced significantly, if we use the matrix inversion lemma to rearrange (6.19) as

$$\check{\mathbf{h}}_t \;=\; \sqrt{\frac{\rho}{M}}\left(\mathbf{I}_M + \frac{\rho}{M}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\mathbf{Q}_t^{-1}\bar{\mathbf{X}}_t^{(L)}\right)^{-1}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\mathbf{Q}_t^{-1}\mathbf{y}_t^{(L)} \tag{6.20}$$

$$\;=\; \sqrt{\frac{\rho}{M}}\Delta_t\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\mathbf{Q}_t^{-1}\mathbf{y}_t^{(L)}. \tag{6.21}$$

This requires trivial inversion of the diagonal matrix $\mathbf{Q}_t^{-1}$, and inversion of the $M\times M$ matrix $\Delta_t = \big(\mathbf{I}_M + \frac{\rho}{M}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\mathbf{Q}_t^{-1}\bar{\mathbf{X}}_t^{(L)}\big)^{-1}$, with typical values of $M = 2$ to 4.

From (6.17) and (6.21), we see that both the variance and the expected value of the soft decisions affect the LMMSE channel estimation. When a symbol decision is less reliable, i.e., $v_{m,t+l}$ is large, the contribution of this symbol to the channel estimate is less. Also, symbols that are far from the center of the sliding window also contribute less to the channel estimation due to the larger value of $\psi_l$.

Using (6.15) and (6.16), the MSE of channel estimates in (6.21) is given by

$$\begin{aligned}
\check{\epsilon}_t^2 \;&=\; \frac{1}{M}\mathrm{Tr}\Big\{\mathbf{I}_M - \check{\mathbf{\Theta}}_t^{\dagger}\check{\mathbf{\Phi}}_t^{-1}\check{\mathbf{\Theta}}_t\Big\} \\
&=\; \frac{1}{M}\mathrm{Tr}\Big\{\mathbf{I}_M - \sqrt{\frac{\rho}{M}}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger}\left(\frac{\rho}{M}\bar{\mathbf{X}}_t^{(L)}\big(\bar{\mathbf{X}}_t^{(L)}\big)^{\dagger} + \mathbf{Q}_t\right)^{-1}\sqrt{\frac{\rho}{M}}\bar{\mathbf{X}}_t^{(L)}\Big\} \\
&=\; \frac{1}{M}\mathrm{Tr}\{\Delta_t\}.
\end{aligned} \tag{6.22}$$

### 6.4.2 Layer 2: Filtering the Coarse Channel Estimates

In the second stage, we design a filter which is applied to the coarse channel estimates to reduce the residual estimation error. Since we assume there is no spatial correlation between the antenna pairs, we apply an independent filter for the gain between each antenna pair. We thus drop the antenna index and design the filter coefficients for an arbitrary antenna pair.

Let $\check{\mathbf{h}}_t^{(K)} = \left[\check{h}_{t-K}, \cdots, \check{h}_t, \cdots, \check{h}_{t+K}\right]^{\mathrm{T}}$. The filtered/improved channel estimate at time $t$ is obtained as

$$\tilde{h}_t = \mathbf{a}_t^{\dagger} \check{\mathbf{h}}_t^{(K)} \tag{6.23}$$

where $\mathbf{a}_t \in \mathcal{C}^{(2K+1)\times 1}$ denotes the coefficients of the filter. By minimizing the cost function $\mathrm{E}\{|h_t - \tilde{h}_t|^2\}$, we obtain the optimum estimator coefficient vector as

$$\mathbf{a}_{\mathrm{o},t} = \boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}^{-1} \boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t} \tag{6.24}$$

where $\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t} = \mathrm{E}\{\check{\mathbf{h}}_t^{(K)} \check{\mathbf{h}}_t^{(K)\dagger}\} \in \mathcal{C}^{(2K+1)\times(2K+1)}$ is the autocorrelation matrix of the coarse channel estimates, and $\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t} = \mathrm{E}\{\check{\mathbf{h}}_t^{(K)} h_t^*\} \in \mathcal{C}^{(2K+1)\times 1}$ is the cross-correlation vector between the coarse channel estimates and the true channel gain at time $t$.

In the past, a number of works, e.g., [38,39], have reported similar filtering of the coarse estimates. To the best of our knowledge, all these works replace the estimates $\check{h}_t$ by the true channel gain $h_t$ to evaluate $\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$ and $\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}$ and accordingly calculate the filter coefficient vector $\mathbf{a}_{\mathrm{o},t}$. We refer to channel estimators built based on this procedure as *mismatched*-DLWF. Obviously, mismatched-DLWF incurs some loss in performance of the receiver. Here, we seek for more accurate estimates of $\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$ and $\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}$ and, hence, a better choice of $\mathbf{a}_{\mathrm{o},t}$. Our analysis presented in the Appendix leads to the following results. The $i$th element of $\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}$ is given by

$$\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}(i) = \gamma_i (1 - \check{\epsilon}_{t+i}^2), \quad \text{for } -K \le i \le K, \tag{6.25}$$

and the $(i,j)$-th element of $\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$ is given by

$$\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}(i,j) = \begin{cases} 1 - \check{\epsilon}_{t+i}^2, & \text{for } -K \le i = j \le K \\ \frac{\gamma_{i-j}}{M} \mathrm{tr}\Big\{ \big(\mathbf{I}_M - \Delta_{t+i}\big)\big(\mathbf{I}_M - \Delta_{t+j}\big) \Big\}, & \text{for } -K \le i \ne j \le K. \end{cases} \tag{6.26}$$

We note that $\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$ is a time-varying matrix and thus it has to be inverted once for each time instant, $t$. To avoid this undesirable complexity, in (6.24), we replace

$\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$ and $\boldsymbol{\theta}_{\check{\mathbf{h}}_t\check{h}_t}$ by their average over the whole packet, denoted by $\bar{\boldsymbol{\Phi}}_{\check{\mathbf{h}}\check{\mathbf{h}}}$ and $\bar{\boldsymbol{\theta}}_{\check{\mathbf{h}}\check{h}}$, respectively. Thus, $\bar{\boldsymbol{\Phi}}_{\check{\mathbf{h}}\check{\mathbf{h}}}$ is inverted only once for the whole packet.

### 6.4.3 Variance of the Channel Estimation Error

The variance of the channel estimation error is required by the MIMO detector in the turbo receiver. In the DLWF channel estimator, we obtained two different channel estimates: the noisy channel estimates in layer one by (6.21) and the filtered channel estimates in layer two by (6.23). The MSE of channel estimation in layer one is given by (6.22), and the MSE of channel in layer two is found by

$$\tilde{\epsilon}_t^2 = 1 - \boldsymbol{\theta}^{\dagger}_{\check{\mathbf{h}}_t\check{h}_t}\boldsymbol{\Phi}^{-1}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}.\boldsymbol{\theta}_{\check{\mathbf{h}}_t\check{h}_t}. \tag{6.27}$$

We also note that $\tilde{\epsilon}_t^2$ is under-estimated because of the independent assumption we made on $w_t'$ in the Appendix. Hence, using this value for the MIMO detector may incur some loss in the performance of the receiver. We empirically found that using the average of $\check{\epsilon}_t^2$ and $\tilde{\epsilon}_t^2$ as the channel estimation error leads to a much better performance of the MIMO detector.

## 6.5 Estimation of the Temporal Correlation $\gamma_\tau$

In Sections 6.3 and 6.4, we developed two channel estimation algorithms for MIMO systems in the context of turbo receivers. For both algorithms, temporal correlation of fading channels $\gamma_\tau$ is assumed to be known at the receiver. However, clearly, this usually is not the case in practice. To present a more realistic measure of performance of the proposed turbo receivers, here, we propose a method of estimating the channel correlation coefficients $\gamma_\tau$ from the channel estimates obtained through pilot symbols.

For transmission, pilot symbols $\mathbf{X}_p \in \mathcal{C}^{T_p \times M}$ are inserted into data symbols with spacing $T_c$. The topic of optimal training to maximize the achievable rate of continuous flat-fading MIMO channels is discussed in [28, 73, 74], and it is beyond the scope of this chapter. We simply take the minimum training length, i.e., $T_p = M$, and the pilot spacing $T_c$ chosen according to the formula

$$T_c = \left\lfloor \frac{0.423}{f_{d,\max} T_s} \right\rfloor \tag{6.28}$$

where $T_s$ is the symbol interval, $f_{d,\max}$ is the maximum Doppler rate of the channel, and $\lfloor \cdot \rfloor$ denotes the floor operator. The pilot symbols are designed to be orthogonal,

i.e., $\mathbf{X}_p^\dagger \mathbf{X}_p = T_p \mathbf{I}_M$. Hence, the channel estimate corresponding to the $l$th pilot symbol is obtained as

$$\begin{aligned} \hat{\mathbf{h}}_{p_l} &= \sqrt{\frac{\rho}{M}} \left( \mathbf{I}_M + \frac{\rho}{M} \mathbf{X}_p^\dagger \mathbf{X}_p \right)^{-1} \mathbf{X}_p \mathbf{y}_{p_l} \\ &= \frac{\sqrt{\rho M}}{M + \rho T_p} \mathbf{X}_p \mathbf{y}_{p_l} \end{aligned} \tag{6.29}$$

where $\mathbf{y}_{p_l}$ denotes the associated received signal.

Next, we obtain an estimate of $\gamma_{T_c}$, denoted by $\hat{\gamma}_{T_c}$, by averaging the correlation between successive channel estimates over all antenna pairs and across time over the received data packet. Using $L_{\text{packet}}$ to denote the packet length, we obtain

$$\hat{\gamma}_{T_c} = \frac{1}{(\lfloor L_{\text{packet}}/T_c \rfloor - 1) MN} \sum_{n=0}^{N-1} \sum_{l=1}^{\lfloor L_{\text{packet}}/T_c \rfloor - 1} \hat{\mathbf{h}}_{n,p_l}^\dagger \hat{\mathbf{h}}_{n,p_{l+1}}. \tag{6.30}$$

The normalized channel fading rate $f_d T_s$ is then obtained as

$$\widehat{f_d T_s} = \frac{1}{2\pi T_c} J_0^{-1}(\hat{\gamma}_{T_c}) \tag{6.31}$$

where $J_0(\cdot)$ is the zero-th order Bessel function of first kind [75]. Once $f_d T_s$ is known, $\gamma_\tau$, for any $\tau$ can be calculated as

$$\gamma_\tau = J_0(2\pi \widehat{f_d T_s} \tau). \tag{6.32}$$

## 6.6 Simulation Results

We present simulation results for a medium MIMO system with $M = N = 3$ and 16 QAM modulation and a large MIMO system, also, with $M = N = 3$ but with 64 QAM modulation. A rate $R = 1/2$ convolutional code of length 8000 with generator polynomials $(117, 155)_{\text{octal}}$ is used. To simulate the time-varying fading channel, we adopt the Jakes' model with the normalized fading rate $f_d T_s = 0.02$. This has the coherence time of approximately 21 symbol intervals. We thus set $T_c$ equal to this value (i.e., $T_c = 21$), for the medium MIMO system that we study. However, our experimental study revealed that $T_c$ should be set somewhat smaller than the coherence time for larger MIMO systems. We set $T_c = 17$ for our large MIMO set-up. We also set $T_p = M$, i.e., to its minimum value. At the initial iteration, channel estimates are obtained from PSAM with interpolation.

We investigate the performance of OWF and DLWF channel estimation algorithms presented in Section 6.3 and Section 6.4 for both the soft-MMSE MIMO detector [16] and the MCMC-MIMO detector [46]. We also present the results of mismatched-DLWF for comparison. The number of iterations between channel estimation and data detection is set equal to 20. We also use the Wiener filter parameters $K = 50$ and $L = 5$ (defined in the previous sections), for both OWF- and DLWF-based receivers. These choices of $K$ and $L$ were found empirically for good performance of both algorithms in our system setup. The bit-error-rate (BER) curves presented below are based on a sufficiently long run of each case, to obtain reliable results. Each point of the curves is finalized after observing at least 1000 frame errors.

### 6.6.1 Receivers with Soft-MMSE Detector

We first show the performance of various receivers with the soft-MMSE MIMO detector. For comparison, we include two standard methods as references: known CSI and PSAM. The BER results for the medium system ($3 \times 3$ MIMO with 16 QAM symbols) are presented in Fig. 6.3. The presented BER curves are those of the OWF channel estimator (Section 6.3), the DLWF channel estimator (Section 6.4), and mismatched-DLWF. Here, we assume that the channel statistic (the Jakes' model parameter $f_d T_s$) is known. The following observations are made.

- The DLWF channel estimator suffers from a minor degradation when compared to the OWF channel estimator. The performance loss is a small fraction of decibel ($\approx 0.15$ dB). We measured the CPU time spent on DLWF and OWF and found that the DLWF channel estimator is over an order of magnitude (about 15 times) less complex than its OWF counterpart.

- The mismatched-DLWF performs about 0.9 dB worse than the DLWF. This justifies the importance of deriving the smoothing filter based on the second-order statistics of the coarse channel estimates.

- At the BER of $10^{-4}$, DLWF performs about 1.5 to 2 dB better than PSAM.

- There is a large gap between OWF/DLWF and the case of known CSI. This is due to the limited performance of the soft-MMSE detector in the presence

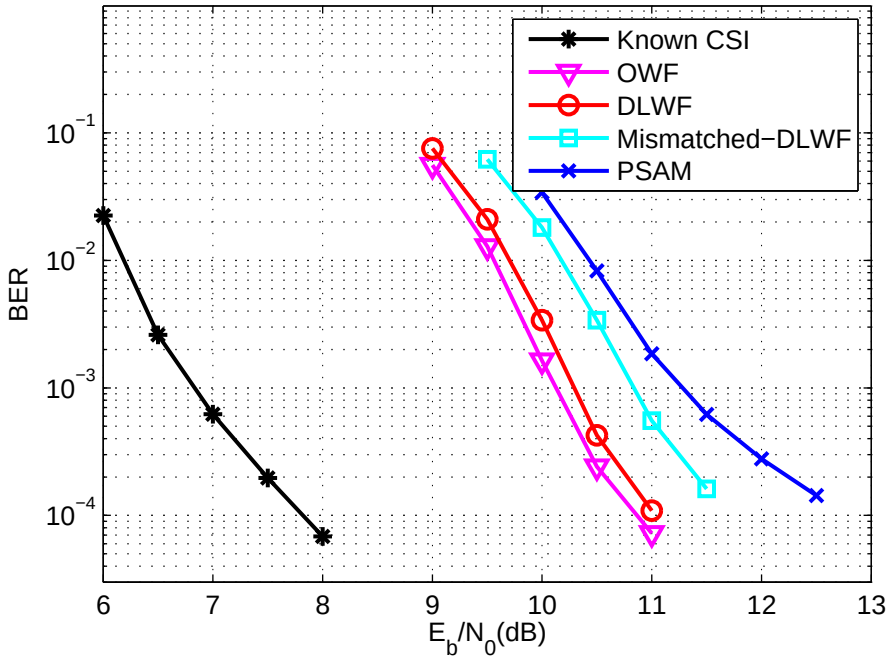


**Figure 6.3**. Performance comparison of a medium MIMO system with soft-MMSE detector. $T_c = 21$. The normalized channel fading rate $f_d T_s$ is known.

of channel estimation error. Later we will show that this gap will be reduced when we replace the soft-MMSE detector by the MCMC detector.

Fig. 6.4 repeats the results of Fig. 6.3 for the case of the large MIMO system. The BER results here are not as good as their counterparts in Fig. 6.3. The difference between the OWF and PSAM, here, is about 1 dB or less. It turns out that the inferior performance of the turbo receiver, when symbol constellations are very large (here, 64 QAM) is related to the disability of the soft-MMSE MIMO detector to resolve the data symbols when channel estimate is inaccurate. It is interesting to note that the mismatched-DLWF performance is inferior to that of PSAM. This we believe results from a combined effect of error propagation and poor performance of the soft-MMSE MIMO detector. Next, we show this problem is resolved if the soft-MMSE MIMO detector is replaced by an MCMC-MIMO detector.

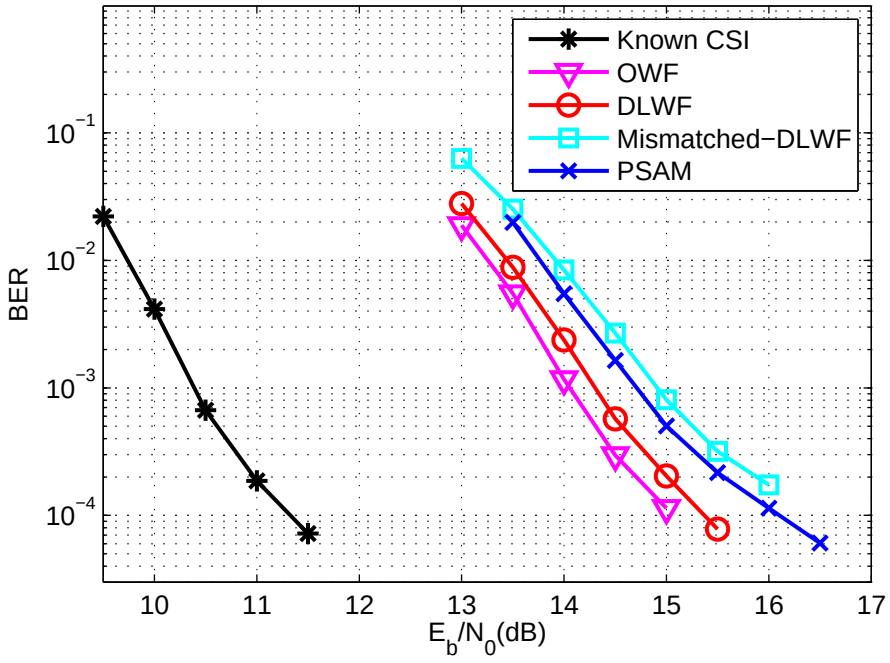


**Figure 6.4**. Performance comparison of a large MIMO system with soft-MMSE detector. $T_c = 17$. The normalized channel fading rate $f_d T_s$ is known.

### 6.6.2 Performance Comparison for MCMC-MIMO Detector

The next set of results that we present follow the same setup as that of Figs. 6.3 and 6.4 with the soft-MMSE MIMO detector replaced by an MCMC-MIMO detector. The MCMC-MIMO detector that we have used here follows [46]. In Fig. 6.5, we use 10 parallel Gibbs samplers and the depth of each Gibbs sampler is also 10. This means we take $10 \times 10 = 100$ samples from the MCMC to calculate the log-likelihood ratio (LLR) values of the information bits. In Fig. 6.6, the number of samples from MCMC is increased to $20 \times 20 = 400$ due to the larger constellation size. The following observations are made.

- Similar to the case of soft-MMSE, the DLWF channel estimator combined with MCMC suffers from a small degradation ($0.2 \sim 0.5$ dB) when compared to the OWF channel estimator. The simulation time for DLWF is again over an order of magnitude less than its OWF counterpart.

- Compared to DLWF, the mismatched-DLWF suffers from a much larger loss

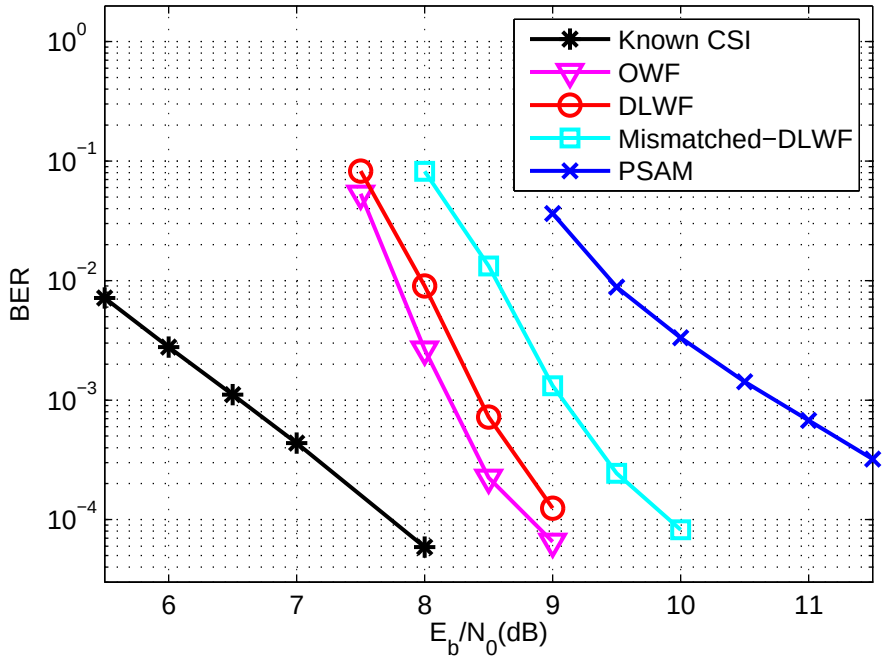


**Figure 6.5**. Performance comparison of a medium MIMO system with MCMC detector. $T_c = 21$. The normalized channel fading rate $f_d T_s$ is known.

(about 1 dB for small system and 2 dB for large system).

- At the BER of $10^{-4}$, the gain brought by the DLWF when compared to PSAM is about 3 dB. This demonstrates that the iterative receiver with the MCMC-MIMO detector efficiently takes the benefit brought by turbo principle.

- The performance gap between the OWF channel estimator and known CSI is less than 1 dB at the BER of $10^{-4}$.

The results that are presented in Figs. 6.5 and 6.6 clearly show the significant receiver gain that is obtained by using the MCMC-MIMO detector instead of the soft-MMSE MIMO detector. This interesting observation reveals that in presence of channel uncertainty, the MCMC-MIMO detector is significantly more robust than the soft-MMSE MIMO detector.

To further compare the impact of replacing the soft-MMSE MIMO detector by its MCMC counterpart, Fig. 6.7 compares the BER curves obtained using both detectors for our large MIMO setup, taken from Figs. 6.4 and 6.6. We have presented the results

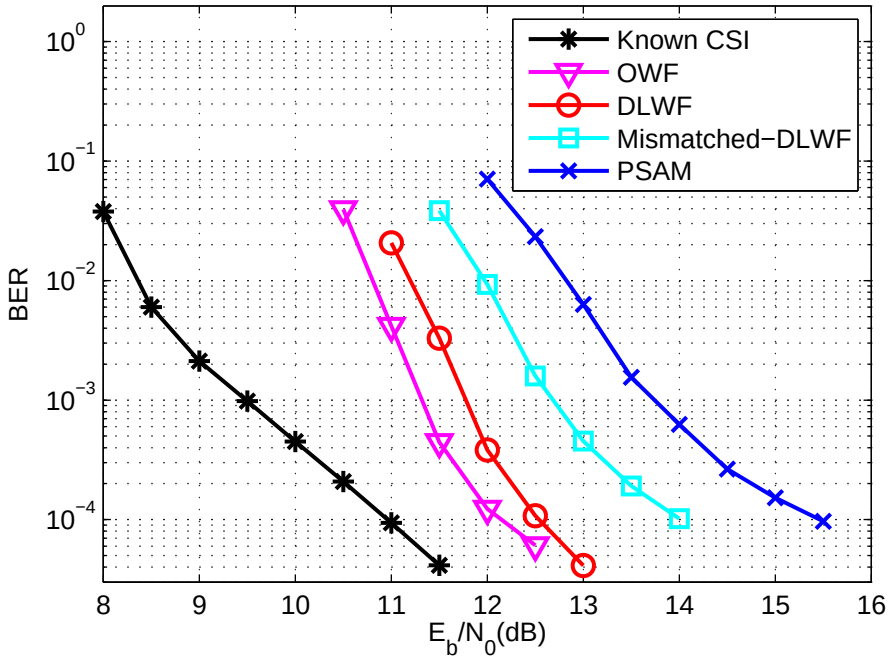


**Figure 6.6**. Performance comparison of a large MIMO system with MCMC detector. $T_c = 17$. The normalized channel fading rate $f_d T_s$ is known.

for the cases of known CSI, PSAM, and the DLWF-based receiver. The cases of known CSI and PSAM are where the channel (estimate) used for detection is fixed (i.e., no channel refinement is made through iterations) and thus, the difference between the two turbo receivers comes from the turbo loop decoding performance only after several iterations. The DLWF-based receiver, on the other hand, refines the channel after each iteration. The following observations are made.

- In general, the MCMC detector outperforms the soft-MMSE detector.
- When CSI is perfectly known, the difference between the system performance resulting from the two MIMO detectors diminishes as SNR increases (equivalently, as BER decreases). However, at lower values of SNR, where BER is high, the MCMC detector outperforms the soft-MMSE detector by a gap of 1 dB or greater.
- In the case of PSAM, when the CSI is imperfect, the system performance resulting from the two MIMO detectors keeps a distance of about 1 dB, for

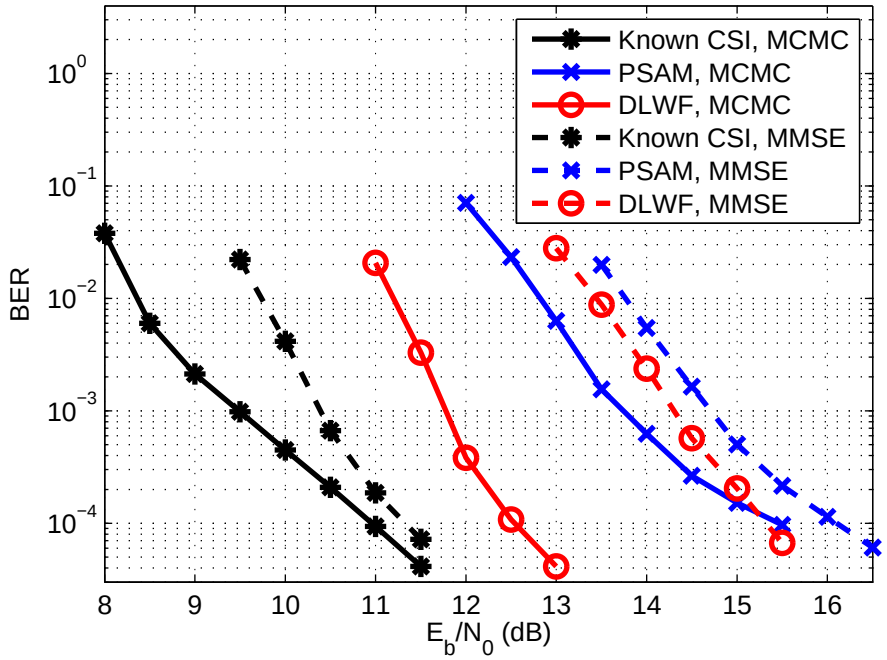


**Figure 6.7**. Performance comparison of soft-MMSE detector and MCMC detector. The normalized channel fading rate $f_dT_s$ is known.

most of the SNR range. From this observation, we infer that, in presence of channel estimation error, a turbo receiver equipped with an MCMC-MIMO detector has a more robust performance than its soft-MMSE counterpart.

- In the case of the DLWF-based receiver (as well as the OWF receiver), where channel estimation is refined at the successive iterations, the MCMC-based receiver outperforms the soft-MMSE-based receiver by a much larger margin. This may be explained as follows. The superior performance of the MCMC-based receivers, over successive iterations, results in a more refined channel estimate, and this in turn improves the overall system performance.

### 6.6.3 Estimation of Temporal Correlation

In Section 6.5, we proposed a method of estimating the temporal correlation coefficients of the Jakes' model of the channel using the channel estimates from the pilot symbols. To verify the effectiveness of the estimated temporal correlation, we let the normalized fading rate vary uniformly in the range $[0.01, 0.02]$, and examine the performance of the proposed turbo receivers with the estimated fading rate and

compare the results with the case where the fading rate is known perfectly. The results for the case of the DLWF channel estimator and the MCMC-MIMO detector, for both the medium and large MIMO systems, are presented in Fig. 6.8. As seen, the difference between the two results is indistinguishable. Only for the large MIMO system, a small difference appears at BER of around $10^{-4}$.

## 6.7 Conclusions

In this chapter, we developed a novel low-complexity turbo receiver for flat fading MIMO channels. Despite significant reduction in complexity, the proposed receiver was found to perform very close to an optimal receiver that uses a data-dependent Wiener filter for channel estimation. The difference in performance is only a fraction of a decibel. We also examined the soft-MMSE and MCMC methods as two possible choices for the MIMO detector part of the receiver. Computer simulations revealed that, in presence of channel estimation error, MCMC performs significantly better than the soft-MMSE detector. Furthermore, assuming a Jakes' model for the channel, we developed a simple method that uses the channel estimates from pilot symbols

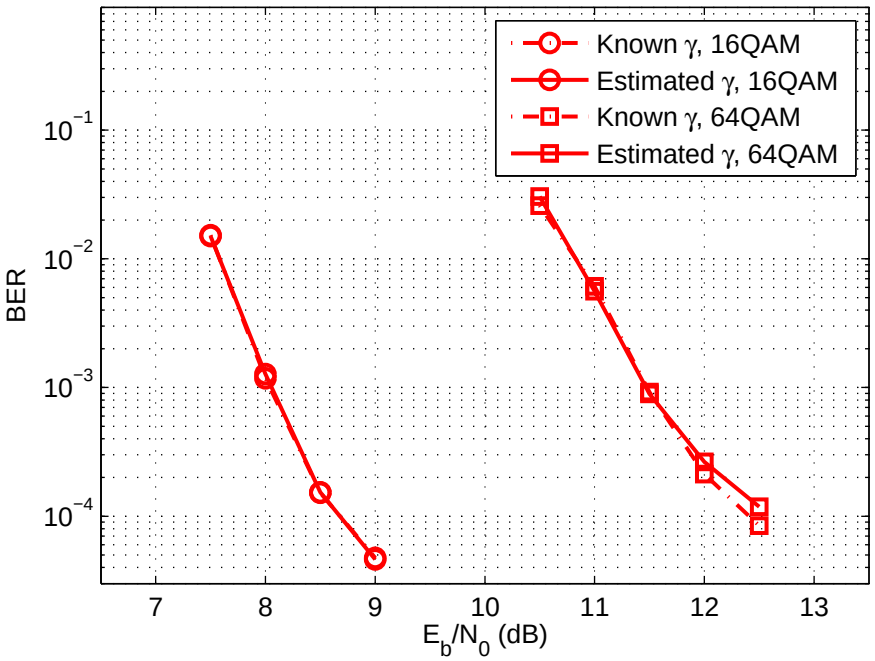


**Figure 6.8**. Performance of DLWF receiver with MCMC detector using estimate correlation coefficients $\gamma_\tau$.

to estimate the fading rate of the channel. Numerical results revealed this simple method is accurate enough to allow the turbo receiver to perform very close to what it would perform when the channel fading rate was known perfectly.

# CHAPTER 7

# CONCLUSION AND FUTURE WORKS

This dissertation proposed a number of iterative receiver designs that address several challenges in MIMO wireless communication systems. The main contributions of the thesis are presented in Chapters 4, 5, and 6.

Chapter 4 took note of the fact that the conventional MCMC-MIMO detector may deteriorate as SNR increases. We suggested and showed through computer simulations that this problem to a great extent can be solved by initializing the MCMC detector with *regulated* states, which are found through linear detectors. We also introduced the novel concept of *staged-MCMC* in a turbo receiver. It starts the detection process at a lower complexity and increases complexity only if the data could not be correctly detected in the present stage of data detection. Computer simulations revealed that this approach could drastically reduce the computational complexity of MCMC detector, yet lead to a satisfactory performance.

Chapter 5 addressed the receiver design for MIMO block fading channels and presented a novel decorrelation receiver to break the correlation between the soft-decision-directed channel estimator (SCE) and MIMO detection. The new design achieves an excellent performance close to that of a genie-aided receiver. We also showed that for fast fading channels, the decorrelation receiver provides a significant gain over the conventional receiver, where no attempt is made to break the correlation between SCE and MIMO detection.

Chapter 6 addressed the joint MIMO detection and channel estimation for time-varying fading channels. To be more realistic, a method of estimating the channel correlation coefficients from the coarse channel estimates obtained through pilot symbols was proposed. Given the estimated channel correlation coefficients, a dual-layer

channel estimator was developed. This method reduces the complexity of the MIMO detector by an order of magnitude at a cost of a negligible degradation in performance, on the order of 0.1 to 0.2 dB.

Furthermore, our study revealed that under the realistic conditions where CSI has to be estimated, hence the available channel estimate will be noisy, the MCMC-MIMO detector outperforms the LMMSE-MIMO detector with a significant margin.

## 7.1 Future Work

### 7.1.1 Receiver Design for MIMO-OFDM

Orthogonal frequency division multiplexing (OFDM) and multiple-input multiple-out (MIMO) techniques have been incorporated in virtually all the current and evolving wireless standards, including WiFi, LTE, and WiMax [76]. OFDM converts a broadband channel into a number of parallel narrowband subcarrier channels with frequency-flat gains, and hence, simplifies the task of channel equalization – each subcarrier channel can be equalized using a single tap equalizer. Moreover, OFDM extension to MIMO channels is a straightforward task, and MIMO-OFDM also inherits the simple equalization/detection benefit of OFDM. The receiver designs proposed in this dissertation have their potential to be applied to MIMO-OFDM systems with appropriate modifications. Such a study should prove very useful for adoption of the methods developed in this thesis to the more practical systems.

### 7.1.2 Receiver Design for Frequency-Selective Fading Channels

Although we developed a number of iterative receivers for MIMO systems in this thesis, the assumption we made is that channel fading is frequency-flat. As discussed in Chapter 2, the channel is frequency-selective when the maximum delay spread is not significantly smaller than the symbol duration. Thus, the channel becomes doubly dispersive. Transmission over doubly dispersive channels suffers from time-varying intersymbol interference (ISI). When the CSI is known and time invariant, the maximum likelihood sequence detector (MLSD) structure based on the Viterbi algorithm (VA) is an optimal detector structure. However, when the channel is unknown and time-varying, MLSD is computationally infeasible. A suboptimal design with joint data detection and channel estimation is another interesting problem to

study.

### 7.1.3 Receiver Design for Underwater Acoustic Channels

Underwater acoustic (UWA) channels are a special form of frequency-selective time-varying channels with very fast fading rates. The methods developed in this thesis are thus a great match to UWA channels and their applications to such channels can be very productive.

# APPENDIX

# DERIVATION OF CORRELATION STATISTICS IN CHAPTER 6

To preserve simplicity of derivations, let us omit the superscript $^{(L)}$ of $\mathbf{y}_t^{(L)}$ and $\mathbf{X}_t^{(L)}$. The elements of correlation vector $\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t} = [\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}(t-K) \cdots \boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}(t) \cdots \boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}(t+K)]^{\mathrm{T}}$ are defined as

$$\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}(i) = \mathrm{E}\big[\check{h}_{t+i} h_t^*\big] = \frac{1}{M}\mathrm{E}\big[\mathbf{h}_t^\dagger \check{\mathbf{h}}_{t+i}\big] = \frac{1}{M}\mathrm{Tr}\Big\{\mathrm{E}\big[\check{\mathbf{h}}_{t+i}\mathbf{h}_t^\dagger\big]\Big\}, \quad \text{for } -K \le i \le K. \tag{A.1}$$

Substituting (6.21) and $\mathbf{y}_{t+i} = \sqrt{\frac{\rho}{M}}\mathbf{X}_{t+i}\mathbf{h}_{t+i} + \mathbf{w}'_{t+i}$ with $\mathbf{w}'_{t+i} = [w'_{t+i-L} \cdots w'_{t+i} \cdots w'_{t+i+L}]^{\mathrm{T}}$ into (A.1), we obtain

$$\begin{aligned}
\boldsymbol{\theta}_{\check{\mathbf{h}}_t h_t}(i) &= \frac{1}{M}\mathrm{Tr}\left\{\mathrm{E}\Big[\sqrt{\frac{\rho}{M}}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger \mathbf{Q}_{t+i}^{-1}\mathbf{y}_{t+i}\mathbf{h}_t^\dagger\Big]\right\} \\
&= \frac{1}{M}\mathrm{Tr}\Big\{\frac{\rho}{M}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger \mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\mathrm{E}\big[\mathbf{h}_{t+i}\mathbf{h}_t^\dagger\big] + \sqrt{\frac{\rho}{M}}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger \mathbf{Q}_{t+i}^{-1}\mathrm{E}\big[\mathbf{w}'_{t+i}\mathbf{h}_t^\dagger\big]\Big\} \\
&= \frac{\gamma_i}{M}\mathrm{Tr}\Big\{\frac{\rho}{M}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger \mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\Big\} = \frac{\gamma_i}{M}\mathrm{Tr}\Big\{\Delta_{t+i}\big(\Delta_{t+i}^{-1} - \mathbf{I}_M\big)\Big\} \\
&= \gamma_i \mathrm{Tr}\Big\{\mathbf{I}_M - \Delta_{t+i}\Big\} = \gamma_i\Big(1 - \frac{1}{M}\mathrm{Tr}\big\{\Delta_{t+i}\big\}\Big) \\
&= \gamma_i\big(1 - \check{\epsilon}_{t+i}^2\big).
\end{aligned} \tag{A.2}$$

Here, $\Delta_{t+i} = \big(\mathbf{I}_M + \frac{\rho}{M}\bar{\mathbf{X}}_{t+i}^\dagger \mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\big)^{-1}$. Note that in (A.2), we assume that the noise $w'_{t+i+l}$ $(-L \le l \le l)$ and $h_t$ are independent.

The elements of the correlation matrix $\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$ are defined as

$$\boldsymbol{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}(i,j) = \mathrm{E}\big[\check{h}_{t+i}\check{h}_{t+j}^*\big] = \frac{1}{M}\mathrm{E}\big[\check{\mathbf{h}}_{t+j}^\dagger \check{\mathbf{h}}_{t+i}\big] = \frac{1}{M}\mathrm{Tr}\Big\{\mathrm{E}\big[\check{\mathbf{h}}_{t+i}\check{\mathbf{h}}_{t+j}^\dagger\big]\Big\}, \tag{A.3}$$

where $-K \le i, j \le K$.

For the diagonal elements of $\mathbf{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}$, i.e., $i=j$, we have $\mathrm{E}\big[\mathbf{y}_{t+i}\mathbf{y}_{t+i}^\dagger\big] = \mathbf{Q}_{t+i} + \frac{\rho}{M}\bar{\mathbf{X}}_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger$. Substituting (6.21) in to (A.3), we have

$$\begin{aligned}
\mathbf{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}(i,i) &= \frac{1}{M}\mathrm{Tr}\bigg\{\sqrt{\frac{\rho}{M}}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\mathrm{E}\big[\mathbf{y}_{t+i}\mathbf{y}_{t+i}^\dagger\big]\Big(\sqrt{\frac{\rho}{M}}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\Big)^\dagger\bigg\} \\
&= \frac{1}{M}\mathrm{Tr}\bigg\{\frac{\rho}{M}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\big(\mathbf{Q}_{t+i}+\frac{\rho}{M}\bar{\mathbf{X}}_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\big)\mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\Delta_{t+i}\bigg\} \\
&= \frac{1}{M}\mathrm{Tr}\bigg\{\frac{\rho}{M}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\Big(\bar{\mathbf{X}}_{t+i}+\frac{\rho}{M}\bar{\mathbf{X}}_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\Big)\Delta_{t+i}\bigg\} \\
&= \frac{1}{M}\mathrm{Tr}\bigg\{\frac{\rho}{M}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\Big(\mathbf{I}_M+\frac{\rho}{M}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\Big)\Delta_{t+i}\bigg\} \\
&= \frac{1}{M}\mathrm{Tr}\Big\{\frac{\rho}{M}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\bar{\mathbf{X}}_{t+i}\Big\} = \frac{1}{M}\mathrm{Tr}\Big\{\Delta_{t+i}\big(\Delta_{t+i}^{-1}-\mathbf{I}_M\big)\Big\} \\
&= 1-\check{\epsilon}_{t+i}^2. \qquad \text{(A.4)}
\end{aligned}$$

For the off-diagonal elements of $\mathbf{\Phi}_{\check{h}_t\check{h}_t}$, i.e., $i\neq j$, we have

$$\begin{aligned}
\mathrm{E}[\mathbf{y}_{t+i}\mathbf{y}_{t+j}^\dagger] &= \mathrm{E}\Big[\big(\sqrt{\frac{\rho}{M}}\mathbf{X}_{t+i}\mathbf{h}_{t+i}+\mathbf{w}'_{t+i}\big)\big(\sqrt{\frac{\rho}{M}}\mathbf{X}_{t+j}\mathbf{h}_{t+j}+\mathbf{w}'_{t+j}\big)^\dagger\Big] \\
&= \frac{\rho\gamma_{i-j}}{M}\mathrm{E}\big[\mathbf{X}_{t+i}\mathbf{X}_{t+j}^\dagger\big] \\
&\approx \frac{\rho\gamma_{i-j}}{M}\bar{\mathbf{X}}_{t+i}\bar{\mathbf{X}}_{t+j}^\dagger. \qquad \text{(A.5)}
\end{aligned}$$

Substituting (6.21) and (A.5) into (A.3), we obtain

$$\begin{aligned}
\mathbf{\Phi}_{\check{\mathbf{h}}_t\check{\mathbf{h}}_t}(i,j) &= \frac{1}{M}\mathrm{Tr}\bigg\{\sqrt{\frac{\rho}{M}}\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+i}^{-1}\mathrm{E}\big[\mathbf{y}_{t+i}\mathbf{y}_{t+j}^\dagger\big]\Big(\sqrt{\frac{\rho}{M}}\Delta_{t+j}\bar{\mathbf{X}}_{t+j}^\dagger\mathbf{Q}_{t+j}^{-1}\Big)^\dagger\bigg\} \\
&= \frac{\gamma_{i-j}}{M}\mathrm{Tr}\bigg\{\Big(\frac{\rho}{M}\Big)^2\Delta_{t+i}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+j}^{-1}\bar{\mathbf{X}}_{t+i}\bar{\mathbf{X}}_{t+j}^\dagger\mathbf{Q}_{t+j}^{-1}\bar{\mathbf{X}}_{t+j}\Delta_{t+j}\bigg\} \\
&= \frac{\gamma_{i-j}}{M}\mathrm{Tr}\bigg\{\Delta_{t+i}\Big(\frac{\rho}{M}\bar{\mathbf{X}}_{t+i}^\dagger\mathbf{Q}_{t+j}^{-1}\bar{\mathbf{X}}_{t+i}\Big)\Big(\frac{\rho}{M}\bar{\mathbf{X}}_{t+j}^\dagger\mathbf{Q}_{t+j}^{-1}\bar{\mathbf{X}}_{t+j}\Big)\Delta_{t+j}\bigg\} \\
&= \frac{\gamma_{i-j}}{M}\mathrm{Tr}\Big\{\Delta_{t+i}\big(\Delta_{t+i}^{-1}-\mathbf{I}_M\big)\big(\Delta_{t+j}^{-1}-\mathbf{I}_M\big)\Delta_{t+j}\Big\} \\
&= \frac{\gamma_{i-j}}{M}\mathrm{Tr}\Big\{\big(\mathbf{I}_M-\Delta_{t+i}\big)\big(\mathbf{I}_M-\Delta_{t+j}\big)\Big\}.
\end{aligned}$$

# REFERENCES


[1] J. H. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Select. Areas Commun.*, vol. 5, no. 5, pp. 871–878, Jun. 1987.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, pp. 585–595, 1999.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[4] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO techniques in WiMAX and LTE: a feature overview," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 86–92, May 2010.

[5] D. Gesbert, M. Shafi, D. shan Shiu, P. J. Smith., and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[6] C. E. Shannon and W. Weaver, "The mathematical theory of communication," *Bell System Bell Labs Technology Journal*, vol. 27, pp. 379–423, 1948.

[7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[8] ——, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[9] L. Zheng and D. D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[10] A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexing in modern MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 186–197, January 2010.

[11] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," *In Proc. 1998 Int. Symp. Sig. Sys. Elect. (ISSSE98)*, pp. 295–300, 1998.

[12] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes (1)," in *IEEE Int. Conf. Commun.*, vol. 2, May 1993, pp. 1064 –1070.

[13] O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Letters*, vol. 4, pp. 161–163, 2000.

[14] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, 2003.

[15] M. Tuchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a *priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673 –683, Mar. 2002.

[16] W. Shin, S. J. Lee, D.-S. Kwon, and J. Kang, "LMMSE channel estimation with soft statistics for Turbo-MIMO receivers," *IEEE Commun. Lett.*, vol. 13, no. 8, pp. 585–87, Aug. 2009.

[17] P. H. Vincent, *An introduction to signal detection and estimation (2nd ed.)*. New York, NY, USA: Springer-Verlag New York, Inc., 1994.

[18] B. Sklar, "Rayleigh fading channels in mobile digital communication systems part i. characterization," *IEEE Commun. Mag.*, vol. 35, no. 7, pp. 90–100, Jul. 1997.

[19] 3GPP, "LTE; volved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception," 3rd Generation Partnership Project (3GPP), TS 132.101 v10.3.0, Jun. 2011.

[20] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall, 2007.

[21] O. Claude and C. Bruno, *MIMO Wireless Communications: From Real-World Propagation to Space-Time Code Design*. Orlando, FL, USA: Academic Press, Inc., 2007.

[22] S. M. Kenneth and A. Mohamed-Slim, *Digital communication over fading channels*. Hoboken, N.J. Wiley-Interscience, 2005.

[23] G. Andrea, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[24] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[25] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Systems Technical Journal*, vol. 47, pp. 957–1000, 1968.

[26] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.

[27] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels [mobile radio]," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[28] Y. Chen and N. Beaulieu, "Optimum pilot symbol assisted modulation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1536–1546, Aug. 2007.

[29] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[30] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[31] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEETrans. Signal. Processing.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[32] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 549–553, Oct. 2002.

[33] M. Loncar, R. Muller, J. Wehinger, C. F. Mecklenbruker, and T. Abe, "Iterative channel estimation and data detection in frequency-selective fading MIMO channels," *Eur. Trans. Telecommun.*, vol. 15, no. 15, pp. 459–470, Sep. 2004.

[34] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 4, pp. 1085–1104, July, 2004.

[35] M. Valenti and B. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1697–1705, Sept. 2001.

[36] N. H. and R. J.A., "Iterative channel estimation and decoding of pilot symbol assisted LDPC coded QAM over flat fading channels," in *Proc. Thirty-Seventh Asilomar Conf. on Signals, Systems and Computers*, vol. 2, Nov. 2003, pp. 2265–2269.

[37] A. Grant, "Joint decoding and channel estimation for linear MIMO channels," in *IEEE Wireless Commun. and Networking Conf.*, vol. 3, Sep. 2000, pp. 1009–1012.

[38] J. Gao and H. Liu, "Decision-directed estimation of MIMO time-varying Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1412–1417, Jul. 2005.

[39] X. Deng, H. A.M., and G.-F. J., "Decision directed iterative channel estimation for MIMO systems," in *IEEE Intl. Conf. on Commun. (ICC)*, vol. 4, May 2003, pp. 2326–2329.

[40] R. Otnes and M. Tuchler, "Iterative channel estimation for Turbo equalization of time-varying frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1918–1923, Nov. 2004.

[41] B. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[42] L. J., K. Pattipati, P. Willett, and G. Levchuk, "Fast optimal and suboptimal any-time algorithms for CDMA multiuser detection based on branch and bound," *IEEE Trans. Commun.*, vol. 52, no. 2, p. 336, Feb. 2004.

[43] R. .P. and C. G., *Monte Carlo Statistical Methods*. Springer-Verlag, 1997.

[44] Y. Z., L. B., and W. X., "Bayesian Monte Carlo multiuser receiver for space-time coded multicarrier CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 19, no. 8, pp. 1625–1637, Aug 2001.

[45] H. Zhu, B. Farhang-Boroujeny, and R.-R. Chen, "On performance of sphere decoding and Markov chain Monte Carlo detection methods," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 669–672, Oct. 2005.

[46] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1896–1909, May 2006.

[47] D. Gamerman and H. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, 1997.

[48] N. Metropolis and S. Ulam, "The Monte Carlo Method," *J. Amer. Stat. Assoc.*, no. 44, pp. 335–341, 1949.

[49] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[50] B. Walsh, "Markov chain Monte Carlo and Gibbs sampling," *Lecture note*, 2004.

[51] A. E. Gelfand 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, 1990.

[52] H. V. Poor and S. Verdu, "Probability of error in MMSE multiuser detection," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 255–270, Feb. 1997.

[53] X. Mao, P. Amini, and B. Farhang-Boroujeny, "Markov chain Monte Carlo MIMO detection methods for high signal-to-noise ratio regimes," in *Proc. IEEE Global Commun. Conf.*, Nov. 2007, pp. 3979 –3983.

[54] S. Laraway and B. Farhang-Boroujeny, "Implementation of a Markov Chain Monte Carlo based multiuser/MIMO detector," vol. 7, Jun. 2006, pp. 3088–3093.

[55] R.-R. Chen, B. Farhang-Boroujeny, and A. Ashikhmin, "Capacity-approaching LDPC codes based on Markov chain Monte Carlo MIMO detection," Jun. 2005, pp. 285–288.

[56] B. Sklar, *Digital Communications: Fundamentals and Applications.* Prentice-Hall, 1988.

[57] Q. Li, C. N. Georghiades, and X. Wang, "An iterative receiver for Turbo-coded pilot-assisted modulation in fading channels," *IEEE Comm. Letters*, vol. 5, no. 4, pp. 145–147, April 2001.

[58] R. Visoz and A. Berthet, "Iterative decoding and channel estimation for space-time BICM over MIMO block fading multipath AWGN channel," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1358 – 1367, aug. 2003.

[59] J. Zheng and B. Rao, "LDPC-coded MIMO systems with unknown block fading channels: soft MIMO detector design, channel estimation, and code optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1504 – 1518, Apr. 2006.

[60] C. Cozzo and B. L. Hughes, "Joint channel estimations and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266–1270, Aug. 2003.

[61] M. Khalighi and J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3165 –3173, Nov. 2006.

[62] R.-R. Chen and R. Peng, "Performance of channel coded noncoherent systems: modulation choice, information rate, and Markov chain Monte Carlo detection," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 2841 –2845, Oct. 2009.

[63] R.-R. Chen, R. Peng, A. Ashikhmin, and B. Farhang-Boroujeny, "Approaching MIMO capacity using bitwise Markov chain Monte Carlo detection," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 423 –428, Feb. 2010.

[64] X. Mao, R.-R. Chen, and B. Farhang-Boroujeny, "Iterative data detection and decoding using list channel estimation and Markov chain Monte Carlo," in *Proc. Int. Symp. Information Theory*, Jun. 2010, pp. 2238 –2242.

[65] B. Lu and X. Wang, "Bayesian blind Turbo receiver for coded OFDM systems with frequency offset and frequency-selective fading," in *Proc. IEEE Int. Conf. Commun. 2002*, vol. 1, pp. 44 –48.

[66] Z. Yang, B. Lu, and X. Wang, "Bayesian Monte Carlo multiuser receiver for space-time coded multicarrier CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1625 –1637, Aug. 2001.

[67] X. Wang and R. Chen, "Blind Turbo equalization in Gaussian and impulsive noise," *IEEE Trans. Veh. Commun.*, vol. 50, no. 4, pp. 1092 –1105, Jul. 2001.

[68] H. Wan, R.-R. Chen, J. W. Choi, A. Singer, J. Preisig, and B. Farhang-Boroujeny, "Markov chain Monte Carlo detection for frequency-selective channels using list channel estimates," *J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1537 –1547, Dec. 2011.

[69] M. Senst and G. Ascheid, "Markov chain Monte Carlo MIMO detection for systems with imperfect channel state information," in *Proc. IEEE Veh. Technol. Conf.*, May 2010, pp. 1 –5.

[70] H. Vikalo, B. Hassibi, and T. Kailath, "Iterative decoding for MIMO channels via modified sphere decoding," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 2299–2311, Nov. 2004.

[71] M. C. Valenti and B. D. Woerner, "Refined channel estimation for coherent detection of Turbo codes over flat-fading channels," *Electron. Lett.*, vol. 34, no. 17, pp. 1648–1649, Aug. 1998.

[72] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory.* Prentice Hall International, 1993.

[73] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. von Helmolt, "Continuous flat-fading MIMO channels: achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1889–1900, Jul. 2005.

[74] S. Ohno and G. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.

[75] J. Proakis, *Digital Communications.* McGraw-Hill Science/Engineering/Math, 2000.

[76] G. L. Stuber, J. R. Barry, T. W. Mclaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.