2017

Channel Estimation and Symbol Detection In Massive MIMO Systems Using Expectation Propagation

Kamran Ghavami
Louisiana State University and Agricultural and Mechanical College, kghava1@lsu.edu

CHANNEL ESTIMATION AND SYMBOL DETECTION IN MASSIVE MIMO SYSTEMS USING EXPECTATION PROPAGATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Electrical Engineering and Computer Sciences

by
Kamran Ghavami
B.S., University of Tehran, 1998
M.S., University of Tehran, 2001
August 2017

To Aram, Janan, and Hafez

Acknowledgments

First and foremost, I would like to express my sincere appreciation to my advisor, Prof. Morteza Naraghi-Pour, for the professional academic guidance he provided to me, and his patience, confidence and care throughout these years. This dissertation would not have been accomplished without his generous support and valuable suggestions.

Also, I would like to express my gratitude to my doctoral committees, Dr. Xuebin Liang, and Dr. Shuangqing Wei for their kind support and thoughtful comments.

I would thank my wife, Aram, who have been always encouraging and supporting me since the early steps toward my doctoral degree.

Table of Contents

List of Tables

List of Figures

Abstract

The advantages envisioned from using large antenna arrays have made massive multiple-input multiple-output systems (also known as massive MIMO) a promising technology for future wireless standards. Despite the advantages that massive MIMO systems provide, increasing the number of antennas introduces new technical challenges that need to be resolved. In particular, symbol detection is one of the key challenges in massive MIMO. Obtaining accurate channel state information (CSI) for the extremely large number of channels involved is a difficult task and consumes significant resources. Therefore for Massive MIMO systems coherent detectors must be able to cope with highly imperfect CSI. More importantly, non-coherent schemes which do not rely on CSI for symbol detection become very attractive.

Expectation propagation (EP) has been recently proposed as a low complexity algorithm for symbol detection in massive MIMO systems , where its performance is evaluated on the premise that perfect channel state information (CSI) is available at the receiver. However, in practical systems, exact CSI is not available due to a variety of reasons including channel estimation errors, quantization errors and aging. In this work we study the performance of EP in the presence of imperfect CSI due to channel estimation errors and show that in this case the EP detector experiences significant performance loss. Moreover, the EP detector shows a higher sensitivity to channel estimation errors in the high signal-to-noise ratio (SNR) regions where the rate of its performance improvement decreases. We investigate this behavior of the EP detector and propose a Modified EP detector for colored noise which utilizes the correlation matrix of the channel estimation error. Simulation results verify that the modified algorithm is robust against imperfect CSI and its performance is significantly improved over the EP algorithm, particularly in the higher SNR regions, and that for the modified detector, the slope of the symbol error rate (SER) vs. SNR plots are similar to the case of perfect CSI.

Next, an algorithm based on expectation propagation is proposed for noncoherent sym-

bol detection in large-scale SIMO systems. It is verified through simulation that in terms of SER, the proposed detector outperforms the pilotbased coherent MMSE detector for blocks as small as two symbols. This makes the proposed detector suitable for fast fading channels with very short coherence times. In addition, the SER performance of this detector converges to that of the optimum ML receiver when the size of the blocks increases. Finally it is shown that for Rician fading channels, knowledge of the fading parameters is not required for achieving the SER gains.

A channel estimation method was recently proposed for multi-cell massive MIMO systems based on the eigenvalue decomposition of the correlation matrix of the received vectors (EVD-based). This algorithm, however, is sensitive to the size of the antenna array as well as the number of samples used in the evaluation of the correlation matrix. As the final work in this dissertation, we present a noncoherent channel estimation and symbol detection scheme for multi-cell massive MIMO systems based on expectation propagation. The proposed algorithm is initialized with the channel estimation result from the EVD-based method. Simulation results show that after a few iterations, the EP-based algorithm significantly outperforms the EVD-based method in both channel estimation and symbol error rate. Moreover, the EP-based algorithm is not sensitive to antenna array size or the inaccuracies of sample correlation matrix.

Chapter 1
Introduction

The limited resources of wireless communication systems, including the limited energy and radio spectrum, is a major bottleneck for serving ever increasing number of users[2, 3] and introducing new wireless services. The capabilities of multiple-antenna systems in improving the system capacity (or throughput), bandwidth efficiency, power efficiency and link reliability of wireless systems was first demonstrated in the vertical Bell laboratories layered space-time (V-BLAST) project [4], as well as in early theoretical studies in [5] and [6]. Since then and in the past two decades, multiple-antenna systems, also known as multiple-input multiple-output (MIMO)[1], have been the subject of intense academic research and have now become an integral part of many standards. Many of the recent wireless standards such as WiFi, WiMAX, HSPA, LTE, etc., rely on MIMO systems.

In general, improvements from MIMO systems are achieved by either combating or exploiting the multipath fading effects of the wireless channels[7]. In spatial diversity techniques, MIMO is used to alleviate the harmful effects of multipath scattering and to increase communication reliability. On the other hand in spatial multiplexing, MIMO is deployed for exploiting the signal scattering of the multipath fading channel to serve a higher number of data streams. All MIMO-based wireless standards use one or both spatial diversity and spatial multiplexing techniques.

The promising advantages of MIMO techniques cannot be realized without the availability of the instantaneous channel coefficients (also known as channel state information (CSI)) either at the receiver or the transmitter. In this regard, MIMO systems can be classified into open-loop or closed-loop systems. In an open-loop MIMO system only the receiver needs the CSI, whereas in a closed-loop system both the receiver and the transmitter use the CSI. In general, in spatial diversity MIMO receivers use CSI for data detection

[1]Following the similar nomenclature of MIMO, the traditional single-antenna systems are also called single-input single-output (SISO).

and in multiplexing techniques MIMO transmitters use CSI for transmit diversity and beamforming. Thus, depending on the availability of CSI at the receiver and/or transmitter, combination of spatial techniques can be used. Table 1.1 shows examples of this combination[7].

Table 1.1: Examples of open-loop and closed-loop MIMO systems.

	Open-loop	**Closed-loop**
Spatial Diversity	Space-Time Coding (STC)	Transmit Selection Diversity (TSD)
Spatial Multiplexing	BLAST	Eigenbeamforming

Conventional MIMO systems, which consist of one multiple-antenna transmitting node and one multiple-antenna receiving node, are referred to as single-user MIMO (SU-MIMO) or point-to-point MIMO. Cellular communication systrems, however, employ multi-user MIMO (MU-MIMO), where one multiple-antenna base station (BS) serves several single-antenna users or mobile stations (MS). Since the users in MU-MIMO are single-antenna systems, their throughput improvement will be limited. However, the entire network will experience increase in the overall throughput[2].

1.1 Massive MIMO

Since the gains offered by MIMO systems scale with the number of transmit and receive antennas, research on high-order MIMO (also referred to as massive MIMO) system has been accelerated in recent years [8, 9, 10, 11]. Early studies have demonstrated the benefits of massive MIMO systems [12], and some field trials have been carried out to show the possibilities and limitations of this technology [13, 14, 15]. Massive MIMO is a MU-MIMO in which the BS is equipped with an order of magnitude larger number of antennas with respect to traditional MIMO systems. For example, while an LTE-A base station can deploy up to 8 antennas, a massive MIMO base station may use tens or even hundreds of antennas. Fig. 1.1 shows a sample prototype of a massive MIMO BS at 3.7 GHz with 160

[2]In fact, all single-antenna nodes can be considered as an integrated node with distributed antennas. Therefore, a MU-MIMO network with single-antenna users may be viewed as a SU-MIMO system.

Figure 1.1: An implemented massive MIMO BS with 160 dual-polarized patch antennas [1].

dual-polarized patch antennas, with the total array size of 60 $\times$ 120 cm [1].

In a MIMO system with N_r and N_t receiving and transmitting antennas, respectively, the small-scale fading channel can be expressed by an $N_r \times N_t$ matrix as H. According to the statistical matrix theory, as the dimensions of this random matrix grow, the distribution of singular values of H become independent of the statistical distribution of its entries and will only depend on the ratio N_t/N_r [16]. An immediate affect of this property is that very tall or very wide (very small or very large N_t/N_r, respectively) channel matrices are vey well conditioned [17]. This property also implies that the histogram of singular values of any single realization of H become very close to the average distribution of singular values. This phenomenon is also known as the channel hardening property. For example, in the reverse link of a MU-MIMO system with a fixed number of mobile users, if N_r increases, the ratio N_t/N_r becomes very small. Based on the channel hardening property, this implies more dominant diagonal and very small off-diagonal entities of $H^H H$, such that the eigenvalues of $H^H H/N_r$ approache to 1 [12]. This property can also be expressed in term of orthogonality of rows or columns of H in two extreme cases of $N_r \ll N_t$ or $N_t \ll N_r$[12]. According to [9], when $N_r \ll N_t$ and $N_t \to \infty$ the row vectors of H will

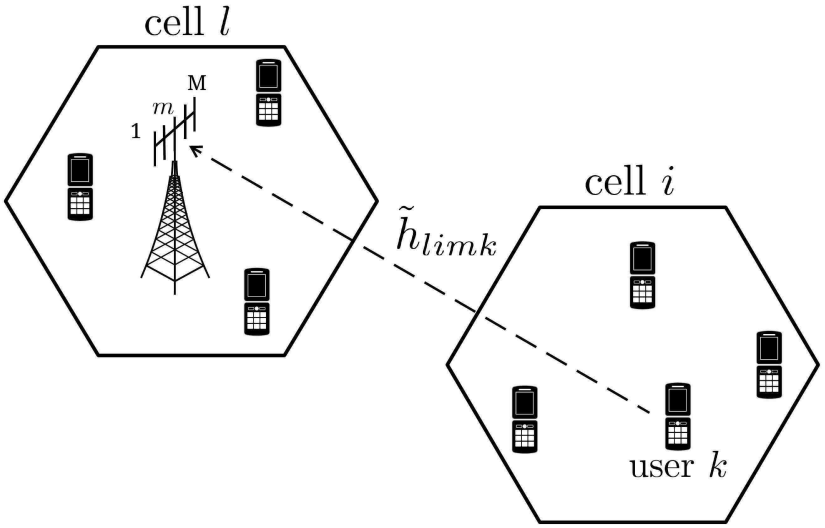


Figure 1.2: Network model in massive MIMO.

become asymptotically orthogonal and hence we have

$$H^H H \approx N_t I_{N_r}, \tag{1.1}$$

and similarly, when $N_t \ll N_r$ and $N_r \to \infty$ the columns of H will become asymptothically orthogonal such that

$$H^H H \approx N_r I_{N_t}. \tag{1.2}$$

Properties (1.1) or (1.2) are also referred to as the favorable propagation conditions.

Consider a MU-MIMO network with L cells with one base station and K users in each cell. Assume the base stations have M antennas and all users are single-antenna terminals. A simple schematic of this network is demonstrated in Fig. 1.2. The narrow band block fading channel is assumed. The channel gain between the m-th antenna of the l-th base station and k-th user located in i-th cell is denoted as $\tilde{h}_{limk}$. Each channel factor $\tilde{h}_{limk}$ can be modeled as follows

$$\tilde{h}_{limk} = \tilde{g}_{limk}\sqrt{\beta_{lik}} \tag{1.3}$$

in which, $\tilde{g}_{limk}$ and β_{lik} represent the independent small-scale fading and large-scale shad-

owing effects, respectively. In this model, $\tilde{g}_{limk}$ is the fast changing fading channel between m-th antenna of the base station l and the k-th user in cell i, and β_{lik} is the slow changing shadowing gain between the l-th base station and the k-th user in cell i. As the indexes show, β_{lik} is independent of the base station's array number m and is identical for all antenna elements at the base station. Denoting $\tilde{\mathbf{g}}_{lik}$ as an $M \times 1$ vector of small-scale fading gains between the l-th base station and k-th user in cell i, the total fading channel between all users of cell i and the l-th base station can be represented by the $M \times K$ matrix $\tilde{G}_{li} = [\tilde{\mathbf{g}}_{li1} \dots \tilde{\mathbf{g}}_{liK}]$. Consequently, by including the shadowing factors, the total channel gain is given as

$$\tilde{H}_{li} = \tilde{G}_{li} D_{li}^{\frac{1}{2}} \tag{1.4}$$

where, D_{li} is a diagonal matrix with β_{lik} values for $k = 1, \dots, K$ on its main diagonal, i.e.

$$D_{li} = \begin{bmatrix} \beta_{li1} & 0 & \dots & 0 \\ 0 & \beta_{li2} & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \beta_{liK} \end{bmatrix} \tag{1.5}$$

To show the effects of deploying a large number of antennas at the base stations, consider the single-cell network, i.e. $L = 1$. To simplify the notations, we ignore the cell indexes. The received vector at the base station in the reverse link is given by the following equation

$$\mathbf{y} = \sqrt{\rho} \tilde{H} \mathbf{x} + \mathbf{n} \tag{1.6}$$

in which, $\mathbf{x}$ is the transmitted vector of unit-energy symbols, $\mathbf{n}$ is the additive zero-mean Gaussian noise with identity covariance matrix, and ρ is the signal power-to-noise power ratio. For large M and based on the favorable channel condition in (1.2), we have

$$\tilde{H}^H \tilde{H} \approx MD \tag{1.7}$$

The capacity for this link is then given as

$$C = \log_2 \det(I_K + \rho H^H H), \tag{1.8}$$

which by using the channel hardening property in (1.7) at large M can be simplified as [8]

$$C \approx \sum_{k=1}^{K} \log_2(1 + M\rho\beta_k) \; bits/s/Hz. \tag{1.9}$$

This formula clearly shows the dependency of the achievable throughput to M. It can be shown that the throughput in (1.9) can be achieved by a simple matched filter (MF) receiver [10].

By assuming time-division duplexing (TDD) mode, the transmissions over forward and reverse links will be done in the same frequency band. Therefore, based on the channel reciprocity property, the forward link's channel matrix will be the transposed version of the reverse link's channel matrix. Consequently, the received vector for the forward link is given as

$$\mathbf{y} = \sqrt{\rho}\tilde{H}^T\mathbf{x} + \mathbf{n} \tag{1.10}$$

The capacity of this link is given by

$$C = \max_{P} \log_2 \det(I_M + \rho\tilde{H}P\tilde{H}^H), \tag{1.11}$$

in which, P is a positive diagonal matrix of allocated powers to transmitting antennas such as $p_1, \ldots, p_K$, where $\sum_{k=1}^{K} p_k = 1$ [18]. Under the favorable channel condition and by using the identity $\det(I + AA^H) = \det(I + A^H A)$, the given capacity simplifies to

$$C \approx \max_{P} \log_2 \det(I_K + \rho MPD) \; bits/s/Hz. \tag{1.12}$$

This equation demonstrates the direct dependency of the throughput to M.

The promising benefits of large antenna arrays in massive MIMO systems can be classified into two major areas of enhancing the channel throughput [8] (as shown in above analysis by (1.9) and (1.12)) and improving the energy efficiency [19]. Despite these advantages, there are some issues in deploying systems with a large number of antennas which need to be addressed. In particular in this report, the implementation issues of channel estimation and symbol detection, as the two dominant challenges in massive MIMO systems, are introduced.

1.2 Challenges in Channel Estimation in Massive MIMO Systems

As discussed previously, both open and closed-loop MIMO configurations require CSI. In MU-MIMO CSI is used for multi-user precoding in the forward link and symbol detection in both forward and reverse links. Due to time-varying nature of cellular channels, CSI changes over time and therefore must be periodically updated at relatively short times[3]. The process of updating or estimating CSI consumes time, bandwidth , power and computational resources. Usually, the CSI estimation is accomplished by transmitting a set of known pilot sequences. The length of pilot should be at least equal to the number of transmitting antennas. Therefore, the required resources for channel estimation in MIMO systems is proportional to the number of transmiting antennas and is independent of the number of receiving antennas.

Since the estimated CSI will be valid for a short time instance, the pilot and payload transmissions should be accomplished in a time/frequency slot in which the channel is nearly constant. This duration depends on several factors, including the carrier frequency, propagation environment, and user mobility, and can be measured as the product of the channel's coherence time (T_c) and coherence bandwidth (B_c). For multi-carrier modulation techniques, such as OFDM, the fading channel will be (frequency) flat and can be assumed

[3]We will see shortly, that the duration in which the channel remains static depends on the coherence parameters of the fading chanel.

static for the duration of the coherence time. For exampel, a channel with $T_c = 1$ ms and $B_c = 100$ kHz will remain constant for about 100 transmission symbols.

Like any wireless technique, MIMO tranceivers can work in either frequency-division duplexing (FDD) or TDD mode. The above fact about the limited number of symbols experiencing a static channel, will differently affect the channel estimation of FDD and TDD modes. If FDD is used, which means the forward and reverse links are in different frequency bands, the downlink and uplink channels (or CSIs) will be different. In the reverse link, the base station receives the pilots transmitted by mobile terminals and estimates their channel. Required resources for uplink channel estimation is independent of the massive number of antennas at the base station. Unlike the reverse link, the forward link CSI must be achieved in two stages. At first the base station transmits a pilot sequence and each user estimates its own downlink CSI. Then terminals must transmit their CSI measurements back to the base station. Since the resources required for this procedure is proportional to the number of the base station's antennas, at best case, the whole $T_c \times B_c$ symbols which are experiencing a static channel, must be dedicated to pilot symbols. Therefor, despite its practical advantages, deploying FDD in massive MIMO is still an open problem [11].

In the case of TDD, in which the forward and reverse transmissions are at the same frequency band but in different time slots, the channel reciprocity can be exploited[4]. Consequently, the estimated uplink CSI can also be used as downlink CSI. Therefore, in TDD mode, first the mobile terminals transmit the pilot sequences and the base station estimates the uplink CSI by receiving those pilots. Next, the base station uses the estimated CSI for detecting uplink symbols as well as downlink beamforming. To achieve the best uplink CSI estimation, the pilot sequences of different mobile users must be orthogonal. However, due to the restriction of the coherence properties of the channel, the number of orthogonal sequences are limited. Accordingly, the pilot sequences employed by adjacent cells may be

[4]While in the TDD the physical channels of forward and reverse links are identical, the two links experience different electrical circuitary. Therefore, pure reciprocity does not exist even in TDD mode. However, by applying proper periodical calibrations, the two links can be approximated as identical[1].

nonorthogonal, leading to the so called pilot contamination problem[8]. We will see shortly that unlike the intera-cell multi-user interference, which can be mitigated be employing large antenna arrays, the inter-cell interference caused by pilot contamination cannot be removed by increasing the number of antennas in base stations.

To show the harmful effects of pilot contamination, we can use the introduced multi-cell massive MIMO model in Fig. 1.2. Assume the worst case scenarion, in which mobile users synchronously and simultaneously transmit pilot sequences for uplink channel estimation. As another important assumption, suppose the complete intera-cell orthogonality among pilots and that the identical set of pilots are used in all cells. By assuming τ as the length of pilots, the pilot sequence of the k-th users in all cell can be considered by a $1 \times \tau$ row vector denoted as $\mathbf{p}_k$. Consequently, we can represent the matrix of all K orthogonal pilots inside each cell by $P = [\mathbf{p}_1^T, \ldots, \mathbf{p}_K^T]^T$, which is a $K \times \tau$ matrix. By intra-cell orthogonality we have $PP^H = \tau I_K$. Without loss of generality, assume the uplink channel estimation in the first base station, i.e. $l = 1$. From (1.6) and by the above assumptions, the received matrix at this base station can be written as

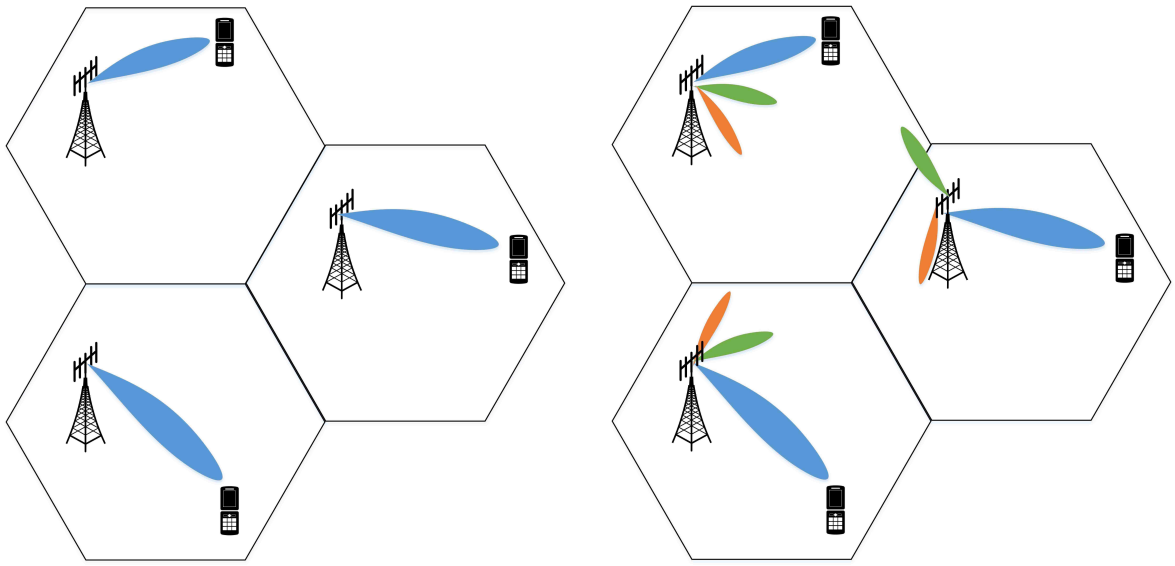
$$Y_1 = \sqrt{\rho} \sum_{l=1}^{L} \tilde{H}_{1l} P + N_1 \tag{1.13}$$

where $\tilde{H}_{1l} \in \mathbb{C}^{M \times K}$ is defined in (1.4), and $N_1 \in \mathbb{C}^{M \times K}$ is the additive noise matrix of the first base station during the pilot transmission.

For channel estimation, the target base station must project the received signals to the space of orthogonal pilots. This can be implemented by multiplying the received matrix by the P^H [10]. Therefore, H_{11} as the estimation of $\tilde{H}_{11}$ is given as

$$H_{11} = \frac{1}{\tau\sqrt{\rho}} Y_1 P^H \tag{1.14}$$

$$= \tilde{H}_{11} + \sum_{i \neq l} \tilde{H}_{il} + \frac{1}{\tau\sqrt{\rho}} N_1 P^H \tag{1.15}$$



(a) (b)

Figure 1.3: Direct inter-cell reverse link interference due to pilot contamination. (a) Perfect beamforming (without pilot contamination), (b) distorted beamforming (with pilot contamination).

The second term in the right hand side of (1.15) shows the effect of pilot contamination which appears as an extra noise to the final estimation. Considering that all variables in this model are complex-valued, reveals that even a small complex noise can adversely affect the phase of resultant estimations. Since the base station uses this estimation for beamforming in forward link, the channel estimation error creates directional side-lobes toward the other users in adjacent cells, which will be an important source of directional interference. While the intra-cell interference can be mitigated by deploying massive MIMO, the harmful effects of this inter-cell interference cannot be alleviated by increasing the number of antennas. This issue is depicted in Fig. 1.3.

In addition to estimation errors, CSI is subject to errors due to aging and quantization [20, 21, 22, 23]. Therefore, assuming perfect CSI in analysis may lead to incorrect or infeasible results.

1.3 Challenges in Symbol Detection in Massive MIMO Systems

Symbol estimation and detection in MIMO is generally a challenging process. To make it more clear, lets consider the reverse link of a MU-MIMO system, as in (1.6). In this model, the base stations receives the $M \times 1$ vector $\mathbf{y}$ and should detect the transmitted symbols in the $K \times 1$ vector $\mathbf{x}$. Assuming $\mathcal{M}$-ary modulation scheme, symbols are selected from the constellation set $\mathcal{A}_{\mathcal{M}}$. If we assume a coherent detection mechanism in which the perfect CSI is available at the basestation, and that transmitted symbols are equally probable, the maximum likelihood (ML) detection rule is given as

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x} \in \mathcal{A}_{\mathcal{M}}^{K}}{\arg\max}\, p(\mathbf{y}|\mathbf{x}, \tilde{H}) \tag{1.16}$$

To solve this optimization problem, the receiver must search $\mathcal{M}^K$ different K-tuples with elements in $\mathcal{A}_{\mathcal{M}}$. Therefore, the optimum MIMO detection by ML is essentially an exhaustive search method which is an NP-hard problem and its complexity increases exponentially with the number of transmitters and the modulation order. Moreover, the performance of linear schemes such as zero-forcing (ZF) and minimum-mean-squared-error (MMSE) decoders (which have polynomial-time complexity [24]) is poor. Successive interference cancellation (SIC) was shown to improve the performance of ZF in the early MIMO project V-BLAST [25] and later was extended to MMSE [26]. However, the performances of MMSE-SIC and ZF-SIC are still far from that of the ML decoder.

Masive MIMO, on the other hand, by employing an order of magnitude more antennas at the base station provides an opportunity for deploying linear detectors. As is shown in [8], under the favorable channel condition, the capacity in (1.9) is achievable by using a simple linear matched filter detector. This fact can also be understood intuitively. Since in favorable channel conditions the channel vectors of different users become mutually orthogonal, the receiver will be able to remove the interference with deploying even a

simple linear algorithm. As explained before, favorable propagation condition will occur in presense of two extreme cases of very large antenna array ($N_r \to \infty$) and very small number of transmitters ($N_t \ll N_r$). However, in many real propagation environments, increasing N_r does not necessarily creates orthogonality [27]. Moreover, to increase the system's spectral efficiency, it is more appealing to serve larger number of users. These facts reveal the possible practical restrictions in using linear detectors in massive MIMO systems. Therefore, nonlinear detection algorithms which provide better performance at the cost of higher complexity, are still possible solutions for symbol detection in massive MIMO systems [10].

1.4 Outline of the Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, we introduce the recently suggested MIMO detection technique based on Expectation Propagation. We will show that the proposed algorithm is very sensitive to the quality of the channel state information at the receiver. Therefore, a modification to the algorithm is suggested to enhance the detector's robustness against the incomplete knowledge of the channel coefficients at the receiver.

A noncoherent detector, based on the Expectation Propagation algorithm, is suggested for Single Input Multiple Output (SIMO) systems in Chapter 3. The inherent phase ambiguity in the estimated channel coefficients can be bypassed by employing differentially encoded modulation symbols. It is shown that the algorithm can easily outperform the coherent Minimum Mean Square Error (MMSE) detectors. Also, the performance of the optimum Maximum Likelihood (ML) detector is achievable with large enough blocks.

In Chapter 4 a joint channel estimation and symbol detection algorithm based on the Expectation Propagation is suggested for multi-user multi-cell MIMO systems. It is shown that by initializing the Expectation Propagation algorithm with results of a rough and inaccurate noncoherent channel estimator, such as the EVD-based algorithm, considerable improvements in channel estimation and symbol detection performances is achieved. It is

also shown that the combination of the two algorithms can significantly decrease the overall rate of erroneous blocks. Finally, the conclusions are given in Chapter 5.

Chapter 2
MIMO DetectionWith Imperfect Channel State Information Using Expectation Propagation

2.1 Introduction

Multiple-input multiple-output (MIMO) technology can significantly increase system capacity (throughput) and improve the reliability of wireless communication systems, and is now incorporated into many wireless standards such as WiFi, WiMAX, LTE, etc. Since the gains offered by MIMO systems scale with the number of transmit and receive antennas, research on high-order MIMO (also referred to as massive MIMO) has been accelerated in recent years [8, 9, 10]. Early studies have demonstrated the benefits of massive MIMO systems [12], and some field trials have been carried out to show the possibilities and limitations of this technology [13, 14, 15].

In massive MIMO systems employing a high-order modulation scheme, symbol detection is a particularly challenging problem. The complexity of the optimal maximum likelihood (ML) decoder is exponential in the number of transmit antennas and it is essentially an exhaustive search method. Moreover, the performance of linear schemes such as zero-forcing (ZF) and minimum-mean-squared-error (MMSE) decoders (which have polynomial-time complexity [24]) is poor. Successive interference cancellation (SIC) was shown to improve the performance of ZF in the early MIMO project V-BLAST [25] and later was extended to MMSE [26]. However, the performances of MMSE-SIC and ZF-SIC are still far from that of the ML decoder.

It is shown in [12] that for a fixed number of transmit antennas N_t[1], as the number of receive antennas N_r increases, the channel vectors become orthogonal. This phenomenon referred to as *channel hardening* occurs when the loading factor $\frac{N_t}{N_r} << 1$. Therefore for

[1]Equivalently, a fixed number of synchronous single-antenna users.

such systems simple linear detectors such as ZF and MMSE detectors show acceptable performance [8]. However, the spectral efficiency of these systems is low due to the small number of transmit antennas N_t. On the other hand, increasing N_t improves the system spectral efficiency, but severely degrades the performance of linear decoders.

Graph-based statistical inference techniques such as Belief Propagation (BP) have proven to be powerful tools for detection problems and also practically viable, particularly in models with a large number of variables or high degrees of freedom [28]. Unfortunately, when the underlying graph has many short cycles, the performance of these algorithms is not satisfactory; and the graph corresponding to symbol detection in MIMO systems is a fully connected graph [29]. To overcome this difficulty, in [29] the authors find a Gaussian Tree Approximation (GTA) on the posterior distribution of the transmitted symbols. The BP algorithm is then used to compute an approximation of this posterior distribution. In [30] GTA has been enhanced with successive interference cancellation (GTA-SIC).

More recently the Expectation Propagation (EP) algorithm of [31] has been applied to symbol detection in MIMO systems [32]. Briefly, EP attempts to find the closest approximation for the conditional marginal distribution of a desired variable in an iterative refinement procedure. Therefore, it can be employed in MIMO detection for finding the posterior distribution of the transmitted symbols. As shown in [32], in terms of symbol error probability, the EP detector outperforms other detectors such as GTA-SIC and MMSE-SIC with low complexity[2].

The performance of EP in [32] is evaluated on the premise that perfect channel state information (CSI) is available at the receiver. However, in MIMO systems, channel coefficients are typically estimated at the receiver from finite-length pilot sequences [33, 34]. In cellular networks using massive MIMO systems, pilot interference from neighboring cells limits the accuracy of channel estimation giving rise to the so-called pilot contamination problem [35]. In addition to estimation errors, CSI is subject to errors due to aging and

[2]For a careful comparison of the computational complexity of the above algorithms we refer the reader to [32].

quantization [20, 36, 22, 23]. In [37] the authors formulate the ML decoder under imperfect CSI and propose recursive tree search algorithms for the implementation of their decoders. Degradation of the performance of ZF in the case of imperfect CSI is analyzed in [23]. However, to the best of our knowledge, the performance of EP algorithm under imperfect CSI has not been studied.

In this paper we show that although channel estimation improves by increasing the signal-to-noise ratio (SNR), surprisingly, at high SNR values, the rate of improvement of symbol error rate (SER) vs. SNR decreases. We investigate this behavior of the EP detector in the case of imperfect CSI and propose a modified detector in order to recover some of the performance loss of the EP detector. Simulation results verify that the proposed modification improves the performance of EP in the case of imperfect CSI, particularly in higher SNR regions, and that for the modified detector the slope of the SER vs. SNR plots are similar to the case of perfect CSI.

The rest of this chapter is organized as follows. The system model is presented in Section 2.2. A brief review on the EP algorithm is presented in Section 2.3. Section 2.4 contains the derivation of EP for the general model with imperfect CSI followed by the calculations of covariance matrix of channel estimation error. Finally the simulation results are presented in Sections 2.5.

Notations: Throughout this paper, small letters (x) are used for scalars, bold small letters ($\mathbf{x}$) for vectors, and capital letters (X) denote matrices. $\mathbb{R}$ and $\mathbb{C}$ represent the set of real and complex numbers, respectively. $\mathfrak{R}(z)$ and $\mathfrak{I}(z)$ denote the real and imaginary parts of the complex variable z. For a set of complex variables $A = \{z_1, z_2, \cdots\}$, we denote $\mathfrak{R}(A) \triangleq \{\mathfrak{R}(z_1), \mathfrak{R}(z_2), \cdots\}$ and $\mathfrak{I}(A) \triangleq \{\mathfrak{I}(z_1), \mathfrak{I}(z_2), \cdots\}$. The superscripts $(.)^T$, $(.)^H$, and $(.)^{-1}$ represent transpose, Hermitian transpose, and matrix inverse, respectively. Also, $\otimes$ denotes the matrix Kronecker product. For a probability density function (PDF) $p(.)$, $\mathbb{E}_p$ denotes the expectation operator with respect to $p(.)$. I_N denotes the $N \times N$ identity matrix. Finally, $\text{vec}(A)$ and $||\mathbf{a}||$ denote the vectorization of the matrix A and the ℓ^2 norm

of vector $\mathbf{a}$, respectively.

2.2 System Model

Consider a MIMO system with N_r and N_t receive and transmit antennas, respectively[3]. The vector of transmitted symbols at each channel use is denoted as $\tilde{\mathbf{u}} = [\tilde{u}_1, \ldots, \tilde{u}_{N_t}]^T \in \mathbb{C}^{N_t \times 1}$, where $\tilde{u}_i$'s are symbols from an M-ary modulation constellation $\tilde{\mathcal{A}}_M$ with average energy E_s. The channel matrix denoted by $\tilde{H}' \in \mathbb{C}^{N_r \times N_t}$ is a realization from a zero-mean complex symmetric Gaussian distribution with covariance matrix $\tilde{R}_h$, i.e., $\tilde{\mathbf{h}}' = \text{vec}(\tilde{H}') \sim \mathcal{CN}(\tilde{\mathbf{h}}'|\mathbf{0}, \tilde{R}_h)$. We assume a block fading channel where the channel matrix $\tilde{H}'$ remains constant for the duration of a transmission block which includes several transmission vectors.

The received vector $\tilde{\mathbf{y}}$ is given by

$$\tilde{\mathbf{y}} = \tilde{H}'\tilde{\mathbf{u}} + \tilde{\mathbf{n}}, \tag{2.1}$$

where $\tilde{\mathbf{y}} \in \mathbb{C}^{N_r \times 1}$, and $\tilde{\mathbf{n}} \in \mathbb{C}^{N_r \times 1}$ is the zero-mean white Gaussian noise vector with $\tilde{\mathbf{n}} \sim \mathcal{CN}(\tilde{\mathbf{n}}|\mathbf{0}, \sigma_n^2 I_{N_r})$. Assuming independent and identically distributed (iid) transmitted symbols, the a posteriori distribution of the transmitted symbols is given by

$$p\left(\tilde{\mathbf{u}}|\tilde{\mathbf{y}}, \tilde{H}'\right) \propto \mathcal{N}\left(\tilde{\mathbf{y}}|\tilde{H}'\tilde{\mathbf{u}}, \sigma_n^2 I_{N_r}\right) \prod_{i=1}^{N_t} \mathbb{I}_{\tilde{u}_i \in \tilde{\mathcal{A}}_M} \tag{2.2}$$

in which $\mathbb{I}_A$ is the indicator function of the event A.

We denote the receiver's estimate of the channel matrix by $\tilde{H}$. Therefore, the receiver's view of the model in (2.1) is given by

$$\tilde{\mathbf{y}} = \tilde{H}\tilde{\mathbf{u}} + \tilde{\mathbf{n}}. \tag{2.3}$$

Consequently, the receiver assumes that the a posteriori distribution of the transmitted

[3]This model is also applicable to a multi-user system in which N_t single-antenna users synchronously transmit to an N_r-antenna receiver.

symbols is given by

$$p\left(\tilde{\mathbf{u}}|\tilde{\mathbf{y}},\tilde{H}\right) \propto \mathcal{N}\left(\tilde{\mathbf{y}}|\tilde{H}\tilde{\mathbf{u}},\sigma_n^2 I_{N_r}\right)\prod_{i=1}^{N_t}\mathbb{I}_{\tilde{u}_i\in\tilde{\mathcal{A}}_M}. \tag{2.4}$$

The distributions in (2.2) and (2.4) have a multiplicative form with respect to the unknown variables which makes it suitable for employing the EP algorithm [31]. However, it should be noted that the receiver assumes the a posterior distribution in (2.4) an it is to this form that the EP algorithm will be applied. In Section 2.4.1 we describe the deleterious consequences of this approach.

2.3 Expectation Propagation

EP is an iterative algorithm for finding the best approximation to a desired distribution from within a tractable family of distributions.

Following the proposed algorithm in [38] and [31], suppose the parameter θ must be estimated from some independent measurements $x_1,\ldots,x_n$. As is common in Bayesian estimation, it is assumed that the prior distribution of θ is known. Therefore the posterior distribution is given by

$$p(\theta|x_1,\ldots,x_n) \propto p(\theta)\prod_{i=1}^{n}p(x_i|\theta) \triangleq \prod_{i=0}^{n}p_i(\theta) \tag{2.5}$$

where $p_0(\theta) \triangleq p(\theta)$ and $p_i(\theta) \triangleq p(x_i|\theta)$ for $i = 1,2\cdots,n$. EP exploits this factorized structure for approximating the above conditional distribution by a distribution from the exponential family, $q(\theta)$, of the form

$$q(\theta) \propto \prod_{i=0}^{n}q_i(\theta) \tag{2.6}$$

where $q_i(\theta)$, $i = 0,1,\cdots,n$ is from an exponential family. Several properties of the exponential family are helpful in simplifying the computations. Two of these properties are extensively used in the computations involved in EP. First is that as in (2.6), multiplication

(or division) of two exponential distributions results in an exponential distribution. Moreover, the parameters of the resulting distribution are easily computed from the parameters of the constituent distributions. Next, the EP algorithm tries to iteratively find the closest $q(\theta)$ to the distribution $p(\theta|x_1, \ldots, x_n)$ where closeness is in terms of the Kullback–Leibler divergence. Therefore, $q(\theta)$ is the solution of the following optimization problem:

$$q^*(\theta) = \arg\min_{q \in \mathcal{F}} KL(p(\theta|x_1, \ldots, x_n) \| q(\theta)) \tag{2.7}$$

where $\mathcal{F}$ is a family of exponential distributions. It turns out that when $\mathcal{F}$ is the exponential family with sufficient statistics $T_1(\theta), T_2(\theta), \ldots, T_S(\theta)$, then the solution of (2.7) is obtained from the *moment matching* condition, namely

$$\mathbb{E}_q[T_i(\theta)] = \mathbb{E}_p[T_i(\theta)], \ i = 1, 2, \ldots, S \tag{2.8}$$

In other words in each step of the optimization we need to match the moments between $q(\theta)$ and $p(\theta|x_1, \cdots, x_n)$. For example if we choose $q(\theta)$ from the family of normal distributions, this is equivalent to equating the mean and variance of $q(\theta)$ and $p(\theta|x_1, \ldots, x_n)$. However, EP implements this process in a subtle way, in which instead of finding the best $q(\theta)$ at once, it finds the best factors of $q(\theta)$ one by one and refines them through successive iterations. At first, the algorithm starts by initializing all the factors $q_i(\theta)$ and consequently $q(\theta)$ itself. Denoting the computed $q(\theta$ at the lth iteration by $q^{(l)}(\theta)$, then all the factors of $q^{(l)}(\theta)$ are updated as follows. To update the i-th factor, a so called *cavity distribution*[4] is derived, in which the effect of the ith factor is eliminated from $q^{(l)}(\theta)$. Therefore, the i-th cavity PDF is given by

$$q^{\setminus i}(\theta) = \frac{q^{(l)}(\theta)}{q_i^{(l)}(\theta)} \tag{2.9}$$

Then by combining $p_i(\theta)$, the i-th factor of $p(\theta|x_1, \ldots, x_n)$, and this cavity factor, a new

[4]Also known as *partial belief*.

intermediate distribution is obtained as

$$\hat{p_i}(\theta) = \frac{1}{Z_i} q^{\setminus i}(\theta) p_i(\theta) \tag{2.10}$$

in which $Z_i = \mathbb{E}_{q^{\setminus i}}[p_i(\theta)] = \int_{-\infty}^{+\infty} p_i(\theta) q^{\setminus i}(\theta)\, d\theta$. Since in general $p_i(\theta)$ and consequently $\hat{p_i}(\theta)$ are not members of the exponential family, the algorithm now finds the closest distribution from the exponential family, $q^{new}(\theta)$, to $\hat{p_i}(\theta)$ using the moment matching condition. After calculating $q^{new}(\theta)$, the refined version of the i-th factor is obtained as

$$q_i^{(l+1)}(\theta) = Z_i \frac{q^{new}(\theta)}{q^{\setminus i}(\theta)} \tag{2.11}$$

After updating all the factors $q_i^{(l+1)}(\theta)$, $i = 0, 1, \cdots, n$, $q^{(l+1)}(\theta)$ is obtained using (2.6), and the process is repeated with the next iteration and until a termination criterion is satisfied. The above procedure is summarized in Algorithm 1. Finally, if we denote the output of the EP algorithm by $\hat{q}(\theta)$, the parameter θ is estimated as $\hat{\theta} = \mathbb{E}_{\hat{q}}[\theta]$.

Data: The main conditional PDF from (2.5)
Result: A member of exponential family as (2.6) which is closest to (2.5)
begin
  Initialize all q_i factors;
  Calculate q by (2.6);
  while *termination criteria has not been met* **do**
    for *i=0,. . . ,n* **do**
      Calculate the cavity PDF by (2.9);
      Calculate the new intermediate PDF by (2.10);
      Find q^{new} by moment matching;
      Update the i-th factor by (2.11);
      Update q by (2.6);
    end
  end
end

Algorithm 1: EP algorithm

2.4 EP algorithm for imperfect CSI

2.4.1 Motivation

In Fig. 2.1 we show the performance of the EP algorithm in terms of symbol error rate (SER) vs. SNR for the cases of perfect and imperfect CSI. For the case of imperfect CSI, the CSI is estimated from a pilot sequence using MMSE estimation. The pilot sequence is assumed to be orthogonal as described in more detail in Section 2.5. The performance of the CSI estimator is also shown in terms of normalized estimation error δ_h vs. SNR, where $\delta_h \triangleq 10\log_{10}||\tilde{\mathbf{h}}' - \tilde{\mathbf{h}}||^2/||\tilde{\mathbf{h}}'||^2$, where as before, $\tilde{\mathbf{h}}' = \text{vec}(\tilde{H}')$ and $\tilde{\mathbf{h}} = \text{vec}(\tilde{H})$. It is assumed that the pilots are transmitted with the same power as the information symbols. Therefore the SNR represents the SNR of the pilot signals as well as the information symbols. Clearly as SNR increases, channel estimation improves and the estimation error is reduced. However, as the figure shows, the performance of EP with the imperfect CSI has a much lower slope and the performance loss with respect to the case of perfect CSI increases with SNR. In particular for SER=10^{-5}, the performance loss is about 15 dB. In this figure (on the right hand side) we also show the channel estimation error δ_h vs. SNR (of the pilot sequence). Therefore this figure can also be viewed as a graph of SER vs. channel estimation error. For example for an SNR of 35 dB, the graph of EP detector with imperfect CSI shows that SER$\approx 10^{-4}$ and the axis on the right shows that the estimation error $\delta_h \approx -39$ dB. Therefore, we conclude that for the estimation error of -39 dB, SER$\approx 10^{-4}$. However, it is important to note here that this value of δ_h is the actual estimation error that a practical system will experience when the pilot symbols have SNR=35 dB and MMSE estimation is used.

To further illustrate the sensitivity of the EP detector to channel estimation errors, the SER performances of MMSE and EP detectors vs. δ_h are compared in Fig. 2.2 for a 20×20 MIMO system. The figure shows that while the EP detector outperforms the MMSE detector for all values of δ_h, its performance is significantly more sensitive to channel estimation error, and that for the EP detector, SER increases sharply with channel estimation error.

Our motivation in this paper is to improve the EP algorithm by incorporating channel estimation into this algorithm so as to recover some of this performance loss.

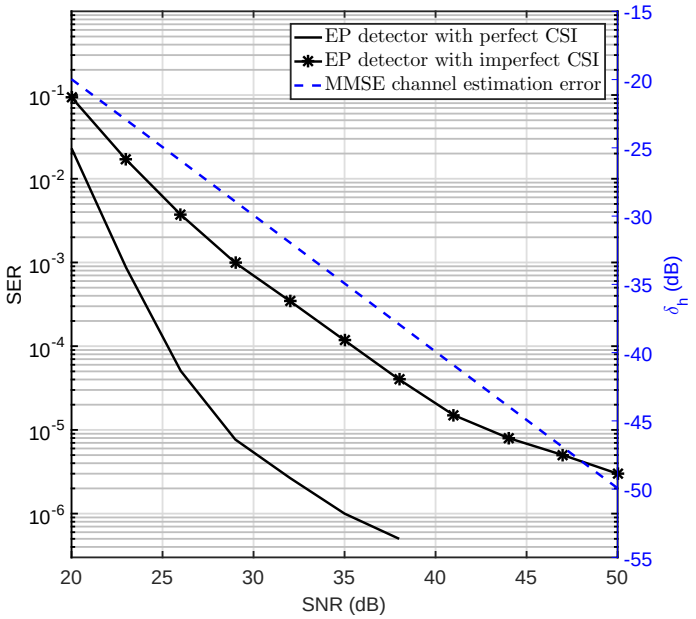


Figure 2.1: Decoding performance of the EP decoder and the channel estimation error for 12×12 antenna configuration and 16-QAM modulation.

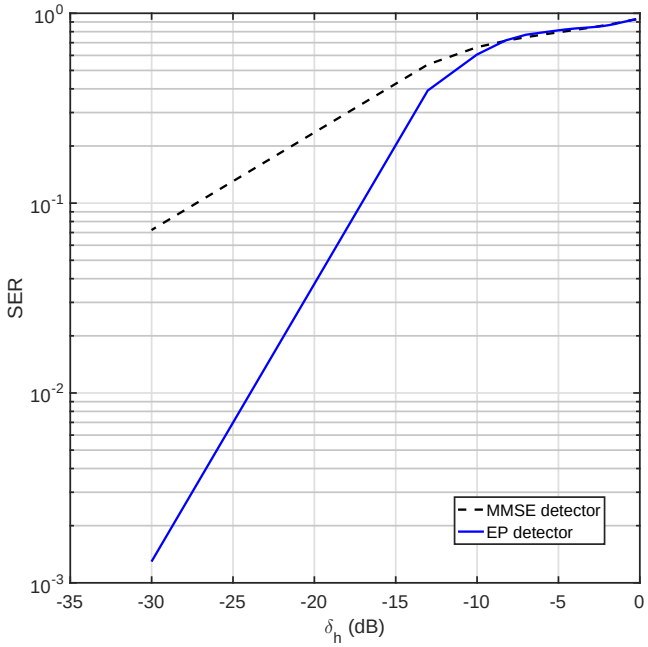


Figure 2.2: Decoding performance of MMSE and EP decoders versus the channel estimation mean square error for 20×20 antenna configuration and 16-QAM modulation.

The reason for this behavior of the EP detector at high SNRs is discussed below. Since

the receiver's view of the a posteriori distribution is given by (2.4), the EP algorithm starts by replacing the non-Gaussian factors in (2.4) by (unnormalized) Gaussian factors to get

$$q(\tilde{\mathbf{u}}) \propto \mathcal{N}\left(\tilde{\mathbf{y}}|\tilde{H}\tilde{\mathbf{u}}, \sigma_n^2 I_{N_r}\right) \prod_{i=1}^{N_t} e^{\gamma_i \tilde{u}_i - \Gamma_i \tilde{u}_i^2/2} \tag{2.12}$$

The constants γ_i and $\Gamma_i > 0$ are then computed iteratively from the EP algorithm [32]. However, as mentioned previously, the true a posteriori distribution is given by (2.2). In other words, the actual mean of the received vector is located at $\tilde{H}'\tilde{\mathbf{u}}$. However, due to the imperfect CSI, the receiver search area is centered around $\tilde{H}\tilde{\mathbf{u}}$, as shown in Fig. 2.3(a). As SNR increases, the distance between the actual mean and the center of the search area, given by $\|(\tilde{H}' - \tilde{H})\tilde{\mathbf{u}}\|$, increases[5]. Equivalently we may consider that the distance remains the same but the search area becomes smaller[6]. This makes it more difficult for EP to find a good approximation to the true a posterior distribution. As a result the performance is degraded. Furthermore, increasing the SNR does not alleviate this problem. Obviously, as SNR goes to infinity, the two SER curves meet. However, as our results show, this does not occur for practical values of SNR.

In order to address this problem and to modify the EP algorithm for the case of imperfect CSI, in the next section we discuss the problem of MIMO channel estimation based on a pilot sequence. We will show that by taking into consideration the channel estimation error, the search area can be aligned toward the actual mean as in Fig. 2.3(b), which leads to a much better detection performance and lower sensitivity of the algorithm to imperfect CSI.

2.4.2 Channel estimation

Due to its simplicity and accuracy, MMSE is commonly used in many pilot-based channel estimation applications. Therefore, we assume that at the receiver CSI is obtained

[5]In practice, an increase in SNR is due to an increase in transmit power E_s.

[6]In other words, we may assume that the transmit power remains constant and the increase in SNR is due to a reduction in noise power.

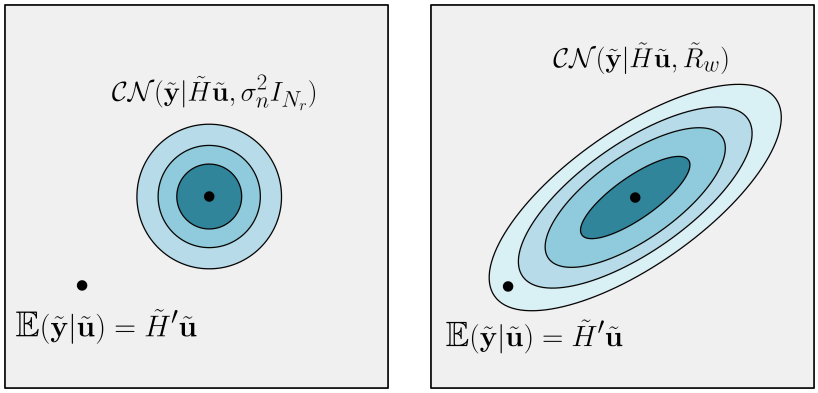


Figure 2.3: The search regions for the EP and Modified EP algorithms. (a) EP with uncorrelated covariance, and (b) Modified EP with correlated covariance.

using the pilot-based MMSE channel estimation algorithm.

Denote the transmitted pilot matrix by $P = [\mathbf{p}(1), \ldots, \mathbf{p}(\tau)] \in \mathbb{C}^{N_t \times \tau}$ where τ denotes the length of the pilot sequence. Therefore according to (2.1), the received matrix is given by $\tilde{Y} = \tilde{H}'P + \tilde{N}$, in which $\tilde{Y} \in \mathbb{C}^{N_r \times \tau}$. Applying matrix vectorization to both sides of this equation results in $\tilde{\mathbf{y}} = \tilde{P}\tilde{\mathbf{h}}' + \tilde{\mathbf{n}}$, where $\tilde{\mathbf{y}} \triangleq \text{vec}(\tilde{Y})$, $\tilde{\mathbf{h}}' \triangleq \text{vec}(\tilde{H}')$, $\tilde{\mathbf{n}} \triangleq \text{vec}(\tilde{N})$, and $\tilde{P} \triangleq P^T \otimes I_{N_r}$. Considering that $\tilde{\mathbf{h}}' \sim \mathcal{CN}(\tilde{\mathbf{h}}'|\mathbf{0}, \tilde{R}_h)$ and $\tilde{\mathbf{n}} \sim \mathcal{CN}(\tilde{\mathbf{n}}|\mathbf{0}, \sigma_n^2 I_{\tau N_r})$, the MMSE estimate of $\tilde{\mathbf{h}}'$ is given by [39]

$$\tilde{\mathbf{h}} = \tilde{R}_h \tilde{P}^H (\tilde{P}\tilde{R}_h\tilde{P}^H + \sigma_n^2 I_{\tau N_r})^{-1}\tilde{\mathbf{y}} \tag{2.13}$$

The receiver uses the estimated channel matrix $\tilde{H}$ instead of the actual channel matrix $\tilde{H}'$. Denoting E as the estimation error, we can write $\tilde{H}' = \tilde{H} + E$, or in the vector form, $\tilde{\mathbf{h}}' = \tilde{\mathbf{h}} + \mathbf{e}$, where $\mathbf{e} = \text{vec}(E)$. Since $\tilde{\mathbf{h}}$ is obtained from a linear operation on $\tilde{\mathbf{y}}$, $\tilde{\mathbf{h}}$ is a multivariate Gaussian vector. Therefore, the estimation error $\mathbf{e}$ is also a zeros-mean Gaussian random vector. From (2.13), The covariance matrix of $\mathbf{e}$ can be calculated as

$$R_e = \tilde{R}_h - \tilde{R}_h \tilde{P}^H (\tilde{P}\tilde{R}_h\tilde{P}^H + \sigma_n^2 I_{\tau N_r})^{-1}\tilde{P}\tilde{R}_h \tag{2.14}$$

Now, if we denote $E = [\mathbf{e}_1, \ldots, \mathbf{e}_{N_t}]$, then $\mathbf{e} = [\mathbf{e}_1^T, \ldots, \mathbf{e}_{N_t}^T]^T$. Therefore, the covariance

matrix of $\mathbf{e}$ can be written as

$$R_e = \begin{bmatrix} R_{1,1} & \dots & R_{1,N_t} \\ \vdots & & \vdots \\ R_{N_t,1} & \dots & R_{N_t,N_t} \end{bmatrix} \tag{2.15}$$

in which $R_{i,j} \triangleq \mathbb{E}[\mathbf{e}_i \mathbf{e}_j^H]$.

It can be verified that $R_E \triangleq \mathbb{E}[EE^H] = \sum_{i=1}^{N_t} R_{i,i}$. Therefore, after calculating R_e from (2.14), R_E can be calculated by adding the N_t matrices of size $N_r \times N_r$ located on the main diagonal of R_e.

From the system model in (2.1), we can write

$$\tilde{\mathbf{y}} = \tilde{H}\tilde{\mathbf{u}} + \tilde{\mathbf{w}} \tag{2.16}$$

where $\tilde{\mathbf{w}} \in \mathbb{C}^{N_r \times 1}$ is a new additive noise vector defined as $\tilde{\mathbf{w}} = E\tilde{\mathbf{u}} + \tilde{\mathbf{n}}$. Assuming a symmetric modulation set such as QAM, for which $\mathbb{E}[\tilde{\mathbf{u}}] = \mathbf{0}$, we get $\mathbb{E}[\tilde{\mathbf{w}}] = \mathbf{0}$ and

$$\tilde{R}_w = E_s R_E + \sigma_n^2 I_{N_r}, \tag{2.17}$$

Considering the fact that the off-diagonal elements of R_E may be non-zero, $\tilde{\mathbf{w}}$ must be treated as colored noise. In the case of perfect CSI, the channel estimation error is zero and, consequently, $\tilde{R}_w = \sigma_n^2 I_{N_r}$. In this case the EP algorithm has very good performance and in fact outperforms other algorithms including GTA [32]. However, in the case of imperfect CSI, R_w is given by (2.17) and as discussed previously, the EP algorithm has a significant performance loss. In the next section we present a Modified EP algorithm for colored noise which is very effective in reducing the sensitivity of the EP algorithm to channel estimation errors.

To simplify notation, we transform (2.16) into an all-real equation as follows

$$\begin{bmatrix} \Re(\tilde{\mathbf{y}}) \\ \Im(\tilde{\mathbf{y}}) \end{bmatrix} = \begin{bmatrix} \Re(\tilde{H}) & -\Im(\tilde{H}) \\ \Im(\tilde{H}) & \Re(\tilde{H}) \end{bmatrix} \begin{bmatrix} \Re(\tilde{\mathbf{u}}) \\ \Im(\tilde{\mathbf{u}}) \end{bmatrix} + \begin{bmatrix} \Re(\tilde{\mathbf{w}}) \\ \Im(\tilde{\mathbf{w}}) \end{bmatrix} \tag{2.18}$$

Equivalently, this can be written as $\mathbf{y} = H\mathbf{u} + \mathbf{w}$ with $\mathbf{y} \in \mathbb{R}^{2N_r \times 1}$, $H \in \mathbb{R}^{2N_r \times 2N_t}$, $\mathbf{u} \in \mathbb{R}^{2N_t \times 1}$, and finally $\mathbf{w} \in \mathbb{R}^{2N_r \times 1}$. In this model, the elements of $\mathbf{u}$ belong to $\mathcal{A}_M = \Re(\tilde{\mathcal{A}}_M) \cup \Im(\tilde{\mathcal{A}}_M)$. In the case of symmetric M-ary QAM, $\mathcal{A}_M$ can be considered as the set of underlying PAM symbols with average energy of $0.5E_s$. Moreover, the covariance matrix of $\mathbf{w}$ is given by

$$R_w = \begin{bmatrix} \frac{1}{2}\tilde{R}_w & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\tilde{R}_w \end{bmatrix} \tag{2.19}$$

Similar to (2.4), the a posteriori PDF for this model is given by

$$p(\mathbf{u}|\mathbf{y}, H) \propto \mathcal{N}(\mathbf{y}|H\mathbf{u}, R_w) \prod_{i=1}^{2N_t} \mathbb{I}_{u_i \in \mathcal{A}_M}. \tag{2.20}$$

In the following section the proposed algorithm in [32] is extended into this more general model.

2.4.3 EP formulation for correlated noise channel

Following the standard methodology of EP algorithm introduced in Section 2.3, the algorithm exploits the factorized form of (2.20) by replacing each factor by a member of the exponential family of distributions. Next, the algorithm refines each factor by applying the moment matching condition. Therefore, we replace each factor $\mathbb{I}_{u_i \in \mathcal{A}_M}$ in (2.20) by a Gaussian PDF $q_i(u_i) = \mathcal{N}(u_i|m_i, \psi_i)$, in which m_i and $\psi_i \geq 0$ are the mean and variance of the distribution, respectively. As a result, the a posteriori PDF in (2.20) can be approximated by $q(\mathbf{u}) \propto \mathcal{N}(\mathbf{y}|H\mathbf{u}, R_w) \prod_{i=1}^{2N_t} q_i(u_i)$, which we write as

$$q(\mathbf{u}) \propto \mathcal{N}(\mathbf{y}|H\mathbf{u}, R_w) \exp(\mathbf{u}^T\mathbf{m} - \frac{1}{2}\mathbf{u}^T V \mathbf{u}), \tag{2.21}$$

where $\mathbf{m} = [m_1/\psi_1, \ldots, m_{2N_t}/\psi_{2N_t}]^T$, and V is a diagonal matrix given by $V = \text{diag}(1/\psi_1, \ldots, 1/\psi_{2N_t})$. Now (2.21) can be written as

$$q(\mathbf{u}) \propto \mathcal{N}(\mathbf{u}|\boldsymbol{\mu}, \Sigma) \tag{2.22}$$

with the covariance matrix Σ and the mean vector $\boldsymbol{\mu}$ where

$$\Sigma = (H^T R_w^{-1} H + V)^{-1} \tag{2.23}$$

and,

$$\boldsymbol{\mu} = \Sigma(H^T R_w^{-1} \mathbf{y} + \mathbf{m}) \tag{2.24}$$

Since the factors $q_i(u_i)$, $i = 1, 2, \cdots, N_t$ depend on distinct variables, they can be updated individually and in parallel by updating the mean and variance pair (m_i, ψ_i) [40]. After updating all the factors, $q(\mathbf{u})$ can be updated from (2.21).

As $q(\mathbf{u})$ is a multivariate Gaussian distribution, its marginal PDFs are also Gaussian. Let $f_i(u_i) = \mathcal{N}(u_i|\mu_i, \nu_i)$ denote the marginal PDF of u_i, where $\mu_i = \boldsymbol{\mu}(i)$ and $\nu_i = \Sigma(i, i)$. EP uses $f_i(u_i)$ for tuning the i-th corresponding factor. Towards this, the i-th cavity distribution is calculated as $q^{\setminus i}(u_i) = f_i(u_i)/q_i(u_i)$. It can be easily shown that $q^{\setminus i}(u_i) \propto \mathcal{N}(u_i|\mu^{\setminus i}, \nu^{\setminus i})$, where

$$\nu^{\setminus i} = (\nu_i^{-1} - \psi_i^{-1})^{-1} \tag{2.25}$$

and

$$\mu^{\setminus i} = \nu^{\setminus i}(\mu_i/\nu_i - m_i/\psi_i). \tag{2.26}$$

Next, by combining the i-th factor from (2.20) and the corresponding cavity distribu-

tion, we compute the following intermediate PDF

$$\hat{p_i}(u_i) = \frac{1}{Z_i} q^{\setminus i}(u_i) \mathbb{I}_{u_i \in \mathcal{A}_M} \tag{2.27}$$

in which $Z_i = \mathbb{E}_{q^{\setminus i}}[\mathbb{I}_{u_i \in \mathcal{A}_M}]$. Now using the moment matching condition, $f_i(u_i)$ is updated by equating its first and second moments with the PDF $\hat{p_i}(u_i)$. Accordingly, the new moments of $f_i(u_i)$ are calculated as $\mu_i^{\text{new}} = \mathbb{E}_{\hat{p_i}}[u_i]$ and $\nu_i^{\text{new}} = \mathbb{E}_{\hat{p_i}}[u_i^2] - (\mu_i^{\text{new}})^2$. It can be verified that

$$Z_i = \sum_{a \in \mathcal{A}_M} q^{\setminus i}(a) \tag{2.28}$$

$$\mathbb{E}_{\hat{p_i}}[u_i] = \frac{1}{Z_i} \sum_{a \in \mathcal{A}_M} a q^{\setminus i}(a) \tag{2.29}$$

$$\mathbb{E}_{\hat{p_i}}[u_i^2] = \frac{1}{Z_i} \sum_{a \in \mathcal{A}_M} a^2 q^{\setminus i}(a) \tag{2.30}$$

After updating $f_i^{\text{new}}(u_i)$, the new $q_i(u_i)$ is obtained from $q_i^{\text{new}}(u_i) \propto f_i^{\text{new}}(u_i)/q^{\setminus i}(u_i)$. Therefore, the values for the new pair $(m_i^{\text{new}}, \psi_i^{\text{new}})$ are given by

$$\psi_i^{\text{new}} = \left((\nu_i^{\text{new}})^{-1} - (\nu^{\setminus i})^{-1}\right)^{-1}, \tag{2.31}$$

and

$$m_i^{\text{new}} = \psi_i^{\text{new}} (\mu_i^{\text{new}} (\nu_i^{\text{new}})^{-1} - \mu^{\setminus i} (\nu^{\setminus i})^{-1}). \tag{2.32}$$

After obtaining all the pairs $(m_i^{\text{new}}, \psi_i^{\text{new}})$, $i = 1, 2, \cdots, 2N_t$, in parallel, the updates Σ^{new} and $\boldsymbol{\mu}^{\text{new}}$ and $q^{\text{new}}(\mathbf{u})$ can be calculated from (2.23), (2.24) and (2.22), respectively. The above procedure is now repeated with this new distribution $q^{\text{new}}(\mathbf{u})$ until a termination criterion is satisfied.

In order to improve the stability of the update rules in (2.31)-(2.32), when $\left((\nu_i^{\text{new}})^{-1} - (\nu^{\setminus i})^{-1}\right)^{-1} \geq$

ϵ, we employ the following smoothing mechanism as suggested in [32], [40]

$$\psi_i^{\text{new}} = \beta \left((\nu_i^{\text{new}})^{-1} - (\nu^{\setminus i})^{-1} \right)^{-1} + (1-\beta)\psi_i \tag{2.33}$$

and,

$$m_i^{\text{new}} = \beta \frac{\mu_i^{new}\nu^{\setminus i} - \mu^{\setminus i}\nu_i^{\text{new}}}{\nu^{\setminus i} - \nu_i^{\text{new}}} + (1-\beta)m_i, \tag{2.34}$$

where $0 < \beta < 1$.

At the end of the iterations, an estimate of the transmitted symbols denoted by $\hat{\mathbf{u}} = [\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_{N_t}]^T$ is obtained as follows[7]. For each $i = 1, 2, \cdots, N_t$,

$$\Re(\hat{u}_i) = \underset{u \in \Re(\tilde{\mathcal{A}}_M)}{\arg\min} |u - \mu(i)|^2 \tag{2.35}$$

$$\Im(\hat{u}_i) = \underset{u \in \Im(\tilde{\mathcal{A}}_M)}{\arg\min} |u - \mu(N_t + i)|^2. \tag{2.36}$$

2.5 Simulation Results and Observations

In this section, we present several simulation results to demonstrate the effectiveness of the proposed modification for the EP detector.

2.5.1 Simulation setup

Our simulation setup is as follows. Given the channel covariance matrix R_h, the simulation starts by generating a channel matrix $\tilde{H}' \in \mathbb{C}^{N_r \times N_t}$ such that $\tilde{\mathbf{h}}' = \text{vec}(\tilde{H}') \sim \mathcal{CN}(\tilde{\mathbf{h}}'|\mathbf{0}, \tilde{R}_h)$. Next, a block of symbols starting with the pilot vectors followed by the information symbols is generated and transmitted over this channel (according to (2.1)) which is assumed to remain constant throughout the transmitted block. The pilots are assumed to be known at the receiver. Using the received signal corresponding to the transmitted pilots, the receiver estimates the channel matrix $\tilde{H}$ using the MMSE algorithm in (2.13). Next, the receiver detects the information symbols using the estimated channel matrix $\tilde{H}$. The SER performance of the receiver is evaluated by repeating this procedure.

[7]Note that for $i = 1, 2, \cdots, N_t$, $\hat{u}_i \in \tilde{\mathcal{A}}_M$.

We have considered two different MIMO channel models. First we assume an uncorrelated channel model where $\tilde{R}_h = I_{N_r N_t}$. Next we consider the Kronecker model where the channel covariance matrix is expressed as the Kronecker product of the covariance matrices of the transmitter side (denoted by $\tilde{R}_t$) and the receiver side (denoted by $\tilde{R}_r$), i.e., $\tilde{R}_h = \tilde{R}_t \otimes \tilde{R}_r$ [41]. For each of the covariance matrices $\tilde{R}_t$ and $\tilde{R}_r$ the following exponential model is considered:

$$\tilde{R}_t = \begin{bmatrix} 1 & \rho_t & \dots & \rho_t^{N_t-1} \\ \rho_t & 1 & \dots & \rho_t^{N_t-2} \\ \vdots & & & \\ \rho_t^{N_t-1} & \rho_t^{N_t-2} & \dots & 1 \end{bmatrix} \tag{2.37}$$

and

$$\tilde{R}_r = \begin{bmatrix} 1 & \rho_r & \dots & \rho_r^{N_r-1} \\ \rho_r & 1 & \dots & \rho_r^{N_r-2} \\ \vdots & & & \\ \rho_r^{N_r-1} & \rho_r^{N_r-2} & \dots & 1 \end{bmatrix}. \tag{2.38}$$

The MMSE channel estimation uses τ pilot symbols where, as before, the pilot matrix is denoted by P. It is shown in [42] that for an uncorrelated channel, the best MMSE estimation is achieved if the orthogonality condition holds for the pilot matrix[8], i.e.,

$$PP^H \propto \tau I_{N_t} \tag{2.39}$$

Moreover, to achieve the maximum channel capacity, the optimal choice for τ is given by $\tau = N_t$ [42]. Therefore in the following simulations we first assume that $PP^H \propto N_t I_{N_t}$. However, in order to investigate the effect of non-orthogonal pilots on the performance of the proposed algorithm, we also present simulation results for semi-orthogonal pilots.

Signal to noise ratio (SNR) is defined as $10\log_{10}(N_t E_s/\sigma_n^2)$[9] and the modulation scheme

[8] We are not aware of an equivalent result for correlated channels.

[9] Note that in the given definition of SNR, the total received power from all transmitters over all the receiving antennas is considered. Therefore, the SNR values are larger than the actual ratio of signal energy

in all the simulations is 16-QAM. The smoothing parameters are chosen as $\beta = 0.2$ and $\epsilon = 5 \times 10^{-7}$ and the Kronecker channel parameters are $\rho_t = 0.1$ and $\rho_r = 0.4$. In all of our simulations each value of SER is evaluated using at least $10^7 N_t$ transmitted symbols.

2.5.2 Numerical results

In the following figures, the EP detector in [32] and the proposed Modified EP detector in this paper are referred to as *EP*, and *Modified EP*, respectively.

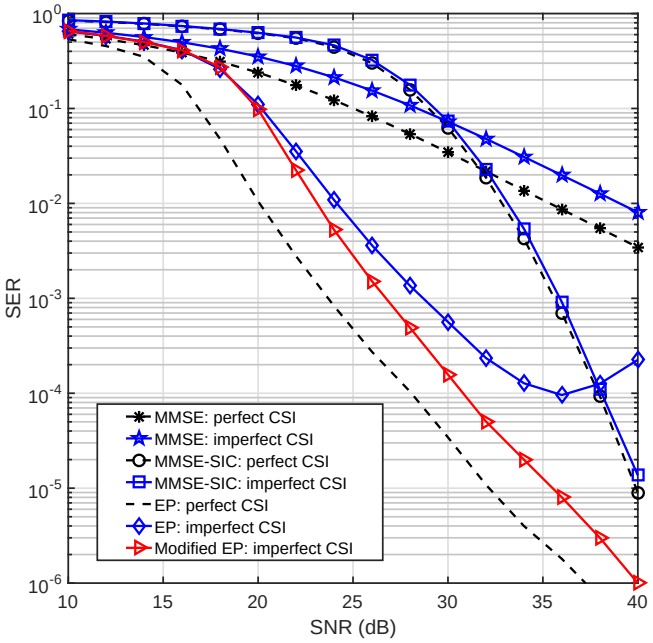


Figure 2.4: Detection performance for the MIMO system with 16-QAM modulation and orthogonal pilot vectors with $N_r = N_t = 32$ and uncorrelated MIMO channel.

Figs. 2.4 and 2.5 show the SER vs SNR for an $N_r \times N_t = 32 \times 32$ and 64×64 MIMO system, respectively, for the MMSE, MMSE-SIC, EP and Modified EP detectors and for the cases of perfect and imperfect CSI. The uncorrelated MIMO channel is assumed. In the case of imperfect CSI, channel estimation is performed using the orthogonal pilot sequences where the SNR of the pilot sequence is the same as the SNR for information symbols shown on the horizontal axis. The EP and Modified EP decoders were run for 10 iterations each. It can be seen that for both antenna configurations, with imperfect CSI, the performance of all detectors is degraded with respect to the case when perfect CSI is available. However,

to the noise power E_s/σ_n^2. In fact, $\text{SNR} = 10\log_{10}(E_s/\sigma_n^2) + 10\log_{10}(N_t)$.

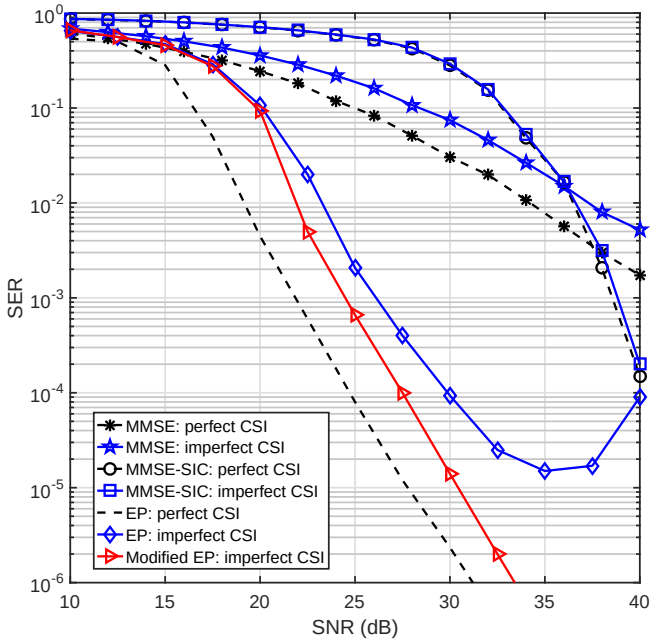


Figure 2.5: Detection performance for the MIMO system with 16-QAM modulation and orthogonal pilot vectors with $N_r = N_t = 64$ and uncorrelated MIMO channel..

the results also show that the EP algorithm is much more sensitive to the CSI errors than MMSE and MMSE-SIC.

Examining the covariance matrix of the total error given in (2.17), we find that at low SNR values the noise plays a dominant role over the CSI estimation errors. Therefore, as Figs 2.4 and 2.5 show, at very low SNR values the SER performances of EP and Modified EP detectors are close. However, as SNR increases, the term in (2.17) related to channel estimation error becomes more dominant. In this case, without compensating for the effect of channel estimation errors, the EP algorithm does not converge to the correct symbol values. This deteriorates the SER performance such that for higher SNR values, the SER curves reach a plateau or even start to rise. On the other hand, these results also show that the slope of the SER vs. SNR plots for the Modified EP detector with imperfect CSI is similar to that of the EP detector with perfect CSI, and that the Modified EP detector provides a much better performance at higher SNR values. For example, Fig. 2.4 shows that for SER of 10^{-4}, the Modified EP detector outperforms the EP detector by about 5 dB. Moreover, at SNR= 36 dB, the performance of EP starts to deteriorate.

As discussed previously, Fig. 2.3(b) provides an intuitive explanation for the performance improvements of the Modified EP detector. As shown in this figure, the Modified EP detector aligns its search area to the direction of the estimation error by employing the proper error covariance matrix R_w and, as a result performs significantly better at higher SNR values.

Figs. 2.6, 2.7 and 2.8 show the SER vs SNR for an $N_r \times N_t = 12 \times 12$, 20×20, and 32×32 MIMO system, respectively, for the MMSE, MMSE-SIC, EP and Modified EP detectors and for the cases of perfect and imperfect CSI, where the Kronecker channel model is assumed. Channel estimation in the case of imperfect CSI is the same as in Figs. 2.4 and 2.5. The EP and Modified EP decoders were run for 10 iterations each. It can be seen that as in the case of uncorrelated channels, the EP algorithm is much more sensitive to CSI errors than MMSE and MMSE-SIC.

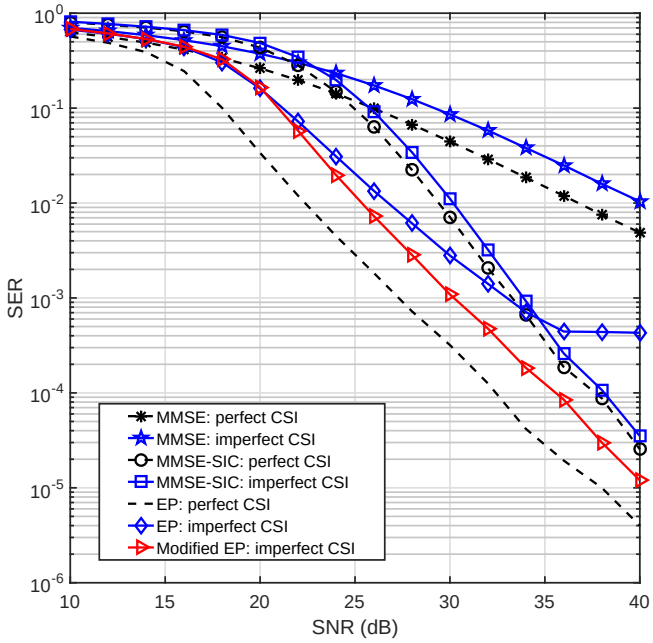


Figure 2.6: Detection performance for the MIMO system with 16-QAM modulation and orthogonal pilot vectors with $N_r = N_t = 12$ over the correlated channel.

Comparing the results in Figs. 2.6-2.8 reveals that the sensitivity of the EP algorithm to channel estimation errors increases for larger antenna arrays. It can be seen that for this correlated channel model, again the proposed modified EP detector helps recover a

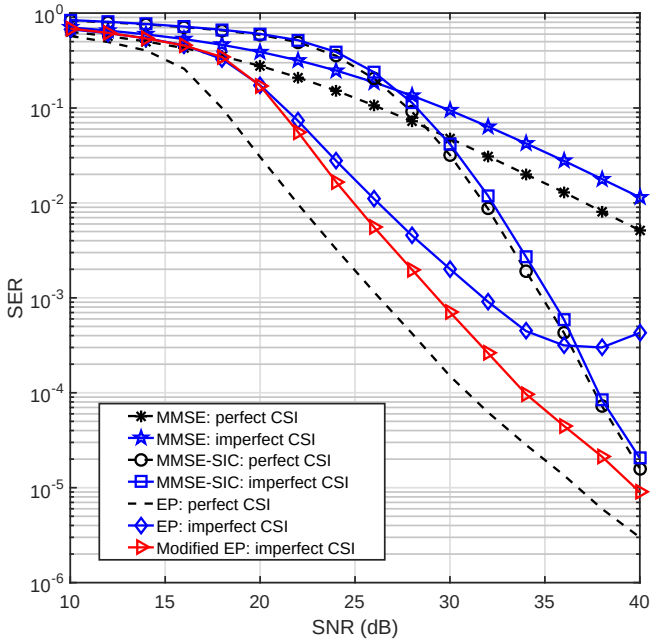


Figure 2.7: Detection performance for the MIMO system with 16-QAM modulation and orthogonal pilot vectors with $N_r = N_t = 20$ over the correlated channel.

great deal of performance loss of the EP detector. For example, Fig. 2.7 shows that for $N_r = N_t = 20$ and SER of 3×10^{-4}, the Modified EP detector outperforms the EP detector by about 5 dB. Moreover, the improvements are larger at higher SNR values.

Fig. 2.9 shows the performances of MMSE, MMSE-SIC, EP and Modified EP in the case of imperfect CSI for an 80×80 MIMO system and the Kronecker channel model. As indicated in the figure, the EP and Modified EP algorithms are evaluated for 2 and 4 iterations. This figure illustrates the effectiveness of the proposed method for large scale systems. For example, for SER of 3×10^{-4}, the proposed Modified EP algorithm outperforms the EP algorithm by more than 3 dB.

It is interesting to note from Fig. 2.9 (as well as in some cases in the previous figures), that the successive interference cancellation technique is not effective and the MMSE decoder outperforms MMSE-SIC decoder. The reason is that for the large value of $N_t = 80$, and the given values of SNR in this figure, the ratio of E_s/σ_n^2 is small (e.g., for SNR=30 dB, $E_s/\sigma_n^2 = 10.97$ dB), and for such small values of E_s/σ_n^2, successive interference cancellation is not effective resulting in poor performance for MMSE-SIC.

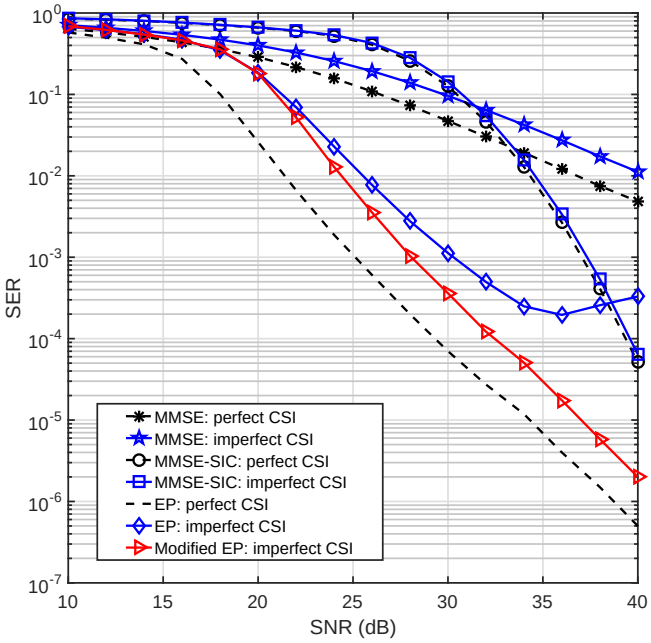


Figure 2.8: Detection performance for the MIMO system with 16-QAM modulation and orthogonal pilot vectors with $N_r = N_t = 32$ over the correlated channel.

As discussed previously, orthogonal pilots result in the best MMSE estimation for uncorrelated channels. However, orthogonal codes are not available for arbitrary values of N_t[10]. Using non-orthogonal pilot sequences increases the channel estimation error. Therefore considering non-orthogonal pilots can also emulate the pilot contamination scenario in massive MIMO systems which inevitably results in increased channel estimation error. For the above reasons, in the next two simulations we assume semi-orthogonal pilot sequences which are generated using Gold sequences [43, 24]. As in the previous simulations, we assume $\tau = N_t$ and the channel model is the same as that described in (2.37)-(2.38). Channel estimation is performed using the pilot sequences as described in Section 2.4.2, and the EP and Modified EP detectors are implemented using the estimated CSI. The EP and Modified EP detectors are simulated with 4 iterations.

Fig. 2.10 shows the SER vs. SNR for the $N_r \times N_t = 32 \times 25$ and 100×25 MIMO system

[10]While OFDM-based wideband MIMO standards, such as LTE use temporal and frequency orthogonality for pilots, here we assume the pilot symbols of all transmit antennas (or users) are transmitted simultaneously in the same frequency band. In other words, orthogonality among the pilot sequences is provided by codeword orthogonality. It is well known that such orthogonal codes (e.g., Walsh-Hadamard codes) do not exist for arbitrary values of N_t [24].

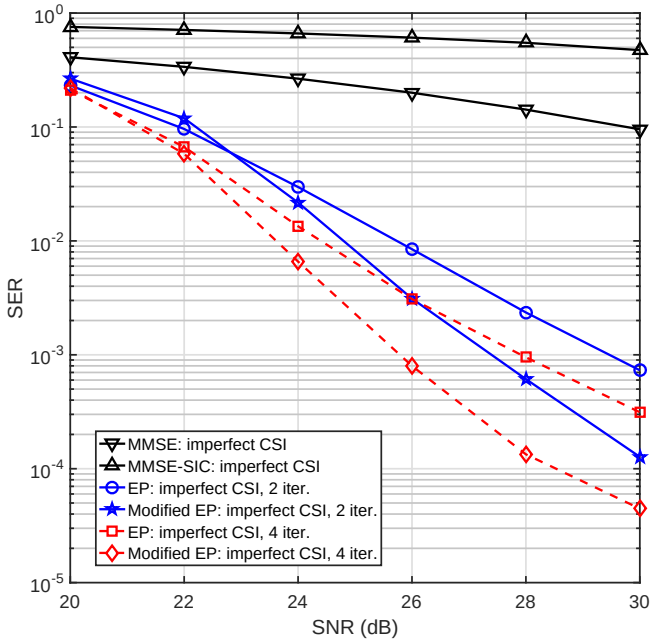


Figure 2.9: Detection performance for the MIMO system with 16-QAM modulation and orthogonal pilot vectors with $N_r = N_t = 80$ over the correlated channel.

configurations. It can be seen that in both configurations, for a large range of SNR values, the EP detector does not converge. After increasing SNR above a certain threshold (about 65 dB for the 32×25 configuration and 55 dB for the 100×25 configuration), the EP detector starts to improve and its SER decreases sharply. In contrast, the Modified EP detector presents a more robust performance against channel estimation errors and outperforms the EP detector for all SNR values. In particular, for SER$= 10^{-4}$, the performance gain of the Modified EP algorithm is more than 20 dB for both configurations. Fig. 2.11 shows a similar result for a 200×30 MIMO system. Similar conclusions can be drawn in this case.

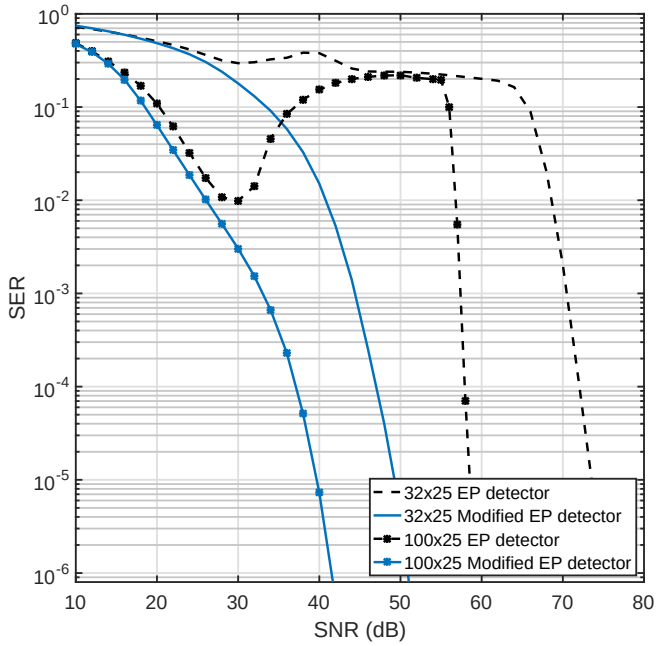


Figure 2.10: Decoding performance of 32×25 and 100×25 MIMO systems, with imperfect CSI, 16-QAM modulations and non-orthogonal pilot vectors .

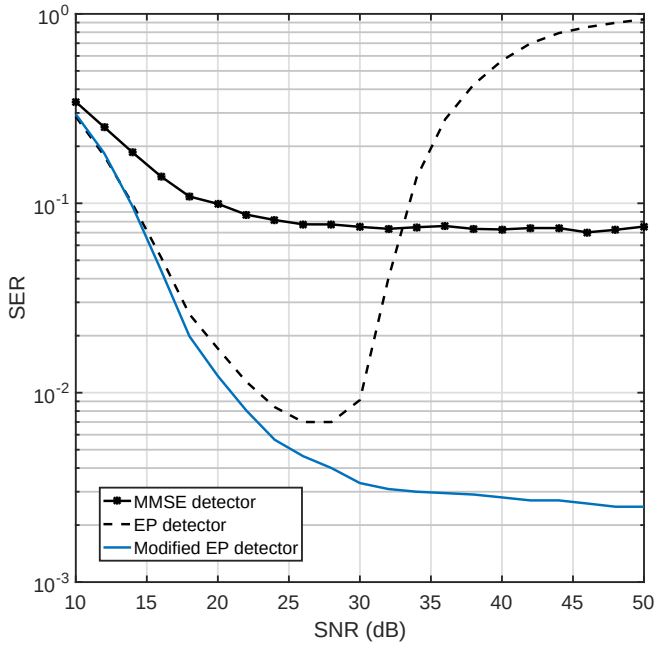


Figure 2.11: Decoding performance of 200×30 MIMO system, with imperfect CSI, 16-QAM modulations and non-orthogonal pilot vectors .

Chapter 3
Noncoherent SIMO Detection by Expectation Propagation

3.1 Introduction

The advantages envisioned from using large antenna arrays have made massive multiple-input multiple-output (MIMO) systems a promising technology for future wireless standards [9, 10]. Massive MIMO systems are expected to provide unprecedented gains in spectral and energy efficiency along with low-complexity linear processing. One of the roadblocks to achieving the promise of massive MIMO systems is the need for accurate knowledge of channel state information (CSI) at the receiver for a large number of channels [44]. Acquiring CSI through transmission of pilot sequences, as is common in many wireless standards, is a resource-consuming task. Since the number of transmitters determines the minimum length of the pilot sequence, increasing the number of antennas in massive MIMO systems intensifies this issue. In particular, the required time for transmitting pilot symbols over the forward link of a massive MIMO system may exceed the nominal coherence time of the channel. Consequently, deployment of massive MIMO systems in frequency division duplexing (FDD) mode is still an open problem [10]. Even in time division duplexing (TDD) mode, due to lack of enough orthogonal sequences, pilots must be shared among cells, giving rise to the so-called pilot contamination problem [9]. This challenge has motivated many researchers to investigate efficient noncoherent or pilotless symbol detection algorithms.

Noncoherent techniques based on sphere decoding determine an efficient search radius in order to reduce the detection complexity [45, 46]. However, sphere decoding based algorithms still suffer from the well-known drawback of high complexity at low SNR values. A noncoherent SIMO detection over uncorrelated Rician fading channel is proposed in [47]. The suggested algorithm is based on a proposed modulation technique in which information

is carried in the power of the transmitted symbols. Since the proposed method relies on averaging the received power over all the receiving antennas, it may need unreasonably large antenna arrays to achieve the desired performance. Therefore, the authors suggest deploying a random code to improve detection performance for smaller antenna arrays [48]. Also, boundary regions in the suggested algorithm are determined by prior knowledge of the channel statistics. As a result, the overall performance strongly depends on the quality of prior knowledge about the Rician channel factor K and the noise power. In [49] the authors propose an optimal constellation for achieving the minimum noncoherent detection error, where the dependency on large scale channel statistics is reduced by applying proper constellation design. Another algorithm in [50] proposes a noncoherent scheme for large scale SIMO systems based on the knowledge of the average received powers from different users. The authors present a specific constellation design such that the receiver can separate the users.

In this paper we propose a noncoherent detection scheme for SIMO systems based on the expectation propagation (EP) algorithm [38, 31]. The proposed EP detector iteratively searches for the best approximation of the joint probability density function (pdf) of the channel coefficients and the transmitted symbols. The output pdf is used for direct estimation of the channel coefficients, as well as the transmitted symbols. We show that for block fading channels, the proposed detector outperforms the pilot-based MMSE detector for both Rayleigh and Rician channels (without prior knowledge of Rician channel factor), for block sizes as small as two symbols. This makes this detector suitable for fast fading channels with very short coherence time. In addition, numerical results show that the performance of this detector converges to that of the optimal maximum likelihood (ML) detector with perfect channel state information (CSI), when the size of the transmitted block increases. Finally, the proposed method does not rely on specific signal constellations and can be used for any differential modulation scheme.

The rest of this chapter is organized as follows. The system model is presented in

Section 3.2. The proposed algorithm is presented in Section 3.3. Finally the simulation results are presented in Sections 3.4.

Notations: Throughout this paper, small letters (x) are used for scalars, bold small letters ($\mathbf{x}$) for vectors, and capital letters (X) for matrices. $\mathbb{R}$ and $\mathbb{C}$ represent the set of real and complex numbers, respectively. The superscripts $(.)^T$, $(.)^H$, and $(.)^{-1}$ represent transpose, Hermitian transpose, and matrix inverse, respectively. Also, $\otimes$ denotes the matrix Kronecker product. For a complex-valued vector $\mathbf{z} = [z_1, z_2, \cdots, z_n]^T$, $\Re(\mathbf{z}) \triangleq [\Re(z_1), \Re(z_2), \cdots, \Re(z_n)]^T$ and $\Im(\mathbf{z}) \triangleq [\Im(z_1), \Im(z_2), \cdots, \Im(z_n)]^T$ where $\Re(z_i)$ and $\Im(z_i)$ denote the real and imaginary parts of the complex variable z_i, respectively. For a pdf $p(.)$, $\mathbb{E}_p$ denotes the expectation operator with respect to $p(.)$. I_N denotes the $N \times N$ identity matrix. Finally, vec(A) and $||\mathbf{a}||$ denote the vectorization of the matrix A and the ℓ^2 norm of vector $\mathbf{a}$, respectively.

3.2 System Model

Consider a SIMO system with N_r receiving antennas. The transmitted symbol at t-th channel use, denoted by $\tilde{s}_t$, belongs to an M-ary modulation constellation $\mathcal{A}_M$ with average energy E_s. The channel vector denoted by $\tilde{\mathbf{h}} \in \mathbb{C}^{N_r \times 1}$ is a circularly symmetric Gaussian random vector with mean vector $\tilde{\mathbf{m}}_0$ and covariance matrix $\tilde{V}_0$, i.e., $\tilde{\mathbf{h}} \sim \mathcal{CN}(\tilde{\mathbf{h}}|\tilde{\mathbf{m}}_0, \tilde{V}_0)$. Note that for modeling the Rician fading channel with factor K, we can choose elements of the mean vector such that $|\tilde{m}_{0i}|^2 = K/(K+1)$, and the covariance matrix as $\tilde{V}_0 = 1/(K+1)I_{N_r}$.

We assume a narrowband block fading channel where $\tilde{\mathbf{h}}$ remains constant for the duration of T transmitted symbols. The t-th received vector $\tilde{\mathbf{y}}_t$ is given by

$$\tilde{\mathbf{y}}_t = s_t\tilde{\mathbf{h}} + \tilde{\mathbf{w}}_t \tag{3.1}$$

where $\tilde{\mathbf{y}}_t \in \mathbb{C}^{N_r \times 1}$, and $\tilde{\mathbf{w}}_t \sim \mathcal{CN}(\tilde{\mathbf{w}}_t|\mathbf{0}, \sigma_w^2 I_{N_r})$. To simplify notation, we consider the

all-real equivalent of (3.1) as follows

$$\begin{bmatrix} \Re(\tilde{\mathbf{y}}_t) \\ \Im(\tilde{\mathbf{y}}_t) \end{bmatrix} = \left(\begin{bmatrix} \Re(s_t) & -\Im(s_t) \\ \Im(s_t) & \Re(s_t) \end{bmatrix} \otimes I_{N_r} \right) \begin{bmatrix} \Re(\tilde{\mathbf{h}}) \\ \Im(\tilde{\mathbf{h}}) \end{bmatrix} + \begin{bmatrix} \Re(\tilde{\mathbf{w}}_t) \\ \Im(\tilde{\mathbf{w}}_t) \end{bmatrix}. \tag{3.2}$$

Equivalently, this can be written as

$$\mathbf{y}_t = S_t \mathbf{h} + \mathbf{w}_t \tag{3.3}$$

with $\mathbf{y}_t \in \mathbb{R}^{2N_r \times 1}$, $\mathbf{h} \in \mathbb{R}^{2N_r \times 1}$, $S_t \in \mathbb{R}^{2N_r \times 2N_r}$, and finally $\mathbf{w}_t \in \mathbb{R}^{2N_r \times 1}$. For the new channel vector $\mathbf{h}$ we have $\mathbf{h} \sim p(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\mathbf{m}_0, V_0)$, where $\mathbf{m}_0 = [\Re(\tilde{\mathbf{m}}_0)^T \ \Im(\tilde{\mathbf{m}}_0)^T]^T$ and $V_0 = \frac{1}{2} I_2 \otimes \tilde{V}_0$. Similarly, $\mathbf{w}_t \sim \mathcal{N}(\mathbf{w}_t|\mathbf{0}, \frac{1}{2}\sigma_w^2 I_{2N_r})$.

Assume a block of T independent transmitted symbols denoted as $\mathbf{s} \triangleq [s_1, \ldots, s_T]$. Using (3.3), the corresponding received vectors are given by $Y \triangleq [\mathbf{y}_1 \ \ldots \ \mathbf{y}_T]$. We are interested in a noncoherent detector where the channel vector $\mathbf{h}$ is unknown and must be estimated along with the transmitted symbols $\mathbf{s}$. The posterior joint distribution of the unknown vectors $\mathbf{s}$ and $\mathbf{h}$ is given by

$$\begin{aligned} p(\mathbf{s}, \mathbf{h}|Y) &\propto p(\mathbf{s}, \mathbf{h}) p(Y|\mathbf{s}, \mathbf{h}) \\ &\propto \left(p(\mathbf{h}) \prod_{t=1}^{T} p(s_t) \right) \prod_{t=1}^{T} p(\mathbf{y}_t|s_t, \mathbf{h}) \end{aligned} \tag{3.4}$$

in which $p(s_t)$ is the probability mass function (pmf) of the t-th transmitted symbol. Since the matrix S_t only depends on the single symbol s_t, we have $p(\mathbf{y}_t|s_t, \mathbf{h}) = p(\mathbf{y}_t|S_t, \mathbf{h}) = \mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{h}, \frac{1}{2}\sigma_w^2 I_{2N_r}\Big)$.

The optimum receiver finds the maximizer of (3.4), i.e.,

$$(\mathbf{s}, \mathbf{h})^* = \underset{\mathbf{s} \in \mathcal{A}_M^T, \mathbf{h} \in \mathbb{R}^{2N_r}}{\arg\max} \ p(\mathbf{s}, \mathbf{h}|Y) \tag{3.5}$$

Due to the complexity of (3.4), finding the optimum solution is generally very difficult and

requires multidimensional integration. The proposed EP algorithm in the next section exploits the multiplicative nature of (3.4) to find a simpler approximation for the conditional joint distribution of $(\mathbf{s}, \mathbf{h})$ such that the marginals can be calculated with much less effort.

3.3 EP formulation for noncoherent detection

Let $\mathcal{F}$ denote a family of exponential distributions. Using EP we exploit the factorized structure of (3.4) to approximate the posterior distribution $p(\mathbf{s}, \mathbf{h}|Y)$ with distributions from $\mathcal{F}$ (for a review of the EP algorithm please refer to [51]). To this end we propose the following approximation for $p(\mathbf{s}, \mathbf{h}|Y)$,

$$q'(\mathbf{s}, \mathbf{h}) \propto \left(p(\mathbf{h}) \prod_{t=1}^{T} p(s_t) \right) q(\mathbf{s}, \mathbf{h}) \tag{3.6}$$

in which $q(\mathbf{s}, \mathbf{h}) \in \mathcal{F}$ may be considered as the approximation to the likelihood function $p(Y|\mathbf{s}, \mathbf{h})$. As such, $q(\mathbf{s}, \mathbf{h})$ can be used for maximum likelihood estimation. However, the receiver can use the prior pdfs to perform a Bayesian estimation from (3.6). Therefore, we only need to apply the EP algorithm to the likelihood pdf of the received vectors. We also note that since the channel vector is continuous and the transmitted symbols are discrete, this will be a hybrid model [51]. The proposed approximation $q(\mathbf{s}, h)$ can be further factorized as

$$q(\mathbf{s}, \mathbf{h}) \propto \prod_{t=1}^{T} q_t(s_t, \mathbf{h}) \propto \prod_{t=1}^{T} q_t(s_t) q_t(\mathbf{h}) \propto q(\mathbf{h}) q(\mathbf{s}) \tag{3.7}$$

in which, $q_t(s_t, \mathbf{h}) \propto q_t(s_t) q_t(\mathbf{h})$, $q(\mathbf{h}) \propto \prod_{t=1}^{T} q_t(\mathbf{h})$ and $q(\mathbf{s}) = \prod_{t=1}^{T} q_t(s_t)$. We note that if we select each factor $q_t(\mathbf{h})$ from $\mathcal{F}$, then $q(\mathbf{h}) \in \mathcal{F}$, and consequently, $q(\mathbf{s}, \mathbf{h}) \in \mathcal{F}$.

We assume that all $q_t(\mathbf{h})$ factors are multivariate Gaussian pdfs, i.e.,

$$q_t(\mathbf{h}) = \mathcal{N}\Big(\mathbf{h}|\mathbf{m}_h(t), V_h(t)\Big). \tag{3.8}$$

Therefore $q(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\mathbf{m}_h, V_h)$, where

$$V_h = \left(\sum_{t=1}^{T} V_h(t)^{-1} \right)^{-1} \tag{3.9}$$

and

$$\mathbf{m}_h = V_h \sum_{t=1}^{T} V_h(t)^{-1} \mathbf{m}_h(t). \tag{3.10}$$

Also, each $q_t(s_t)$ factor is assumed to be the pmf of its corresponding random symbol s_t. Considering the M-ary modulation constellation $\mathcal{A}_M = \{a_1, \ldots, a_M\}$, the pmf for s_t is the set of probabilities as follows

$$q_t(s_t) = \Big\{ \mathbb{P}[s_t = a_1], \ldots, \mathbb{P}[s_t = a_M] \Big\}. \tag{3.11}$$

We use the t-th factor in the likelihood function, i.e., $p(s_t, \mathbf{h}|\mathbf{y}_t)$, for refining the t-th approximating distribution, $q_t(s_t, \mathbf{h})$. Toward this, the t-th cavity distribution is calculated as $q^{\setminus t}(\mathbf{s}, \mathbf{h}) \propto q(\mathbf{s}, \mathbf{h})/q_t(s_t, \mathbf{h}) \propto q^{\setminus t}(\mathbf{h}) q^{\setminus t}(\mathbf{s})$, where $q^{\setminus t}(\mathbf{h}) \propto q(\mathbf{h})/q_t(\mathbf{h})$ and $q^{\setminus t}(\mathbf{s}) = q(\mathbf{s})/q_t(s_t) = \prod_{i=1, i\neq t}^{T} q_i(s_i)$. It can be shown that $q^{\setminus t}(\mathbf{h}) \propto \mathcal{N}(\mathbf{h}|\mathbf{m}_h^{\setminus t}, V_h^{\setminus t})$, where

$$V_h^{\setminus t} = (V_h^{-1} - V_h(t)^{-1})^{-1} \tag{3.12}$$

$$\mathbf{m}_h^{\setminus t} = V_h^{\setminus t}(V_h^{-1}\mathbf{m}_h - V_h(t)^{-1}\mathbf{m}_h(t)) \tag{3.13}$$

Next, by combining the t-th factor of $p(Y|\mathbf{s}, \mathbf{h})$, given in (3.4), and the corresponding cavity distribution, we compute the following intermediate pdf

$$\hat{p}_t(\mathbf{s}, \mathbf{h}) = \frac{1}{Z_t} q^{\setminus t}(\mathbf{s}, \mathbf{h}) p(\mathbf{y}_t|s_t, \mathbf{h}), \tag{3.14}$$

where $Z_t = \mathbb{E}_{q^{\setminus t}(\mathbf{s},\mathbf{h})}[p(\mathbf{y}_t|s_t,\mathbf{h})]$ is the normalization constant and can be calculated as

$$Z_t = \sum_{s_t \in \mathcal{A}_M} \mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, S_t V_h^{\setminus t} S_t^T + \frac{1}{2}\sigma_w^2 I_{2N_r}\Big) \tag{3.15}$$

Since $\hat{p}_t(\mathbf{s},\mathbf{h})$ is not a member of the exponential family, it is replaced with the closest distribution in $\mathcal{F}$, denoted as $q^{new}(\mathbf{s},\mathbf{h}) = q^{new}(\mathbf{h})q^{new}(\mathbf{s})$, where closeness is in the sense of Kullback-Leibler (K-L) divergence. Therefore,

$$q^{new}(\mathbf{s},\mathbf{h}) = \underset{q(\mathbf{s},\mathbf{h}) \in \ \mathcal{F}}{\arg\min}\, \mathrm{KL}\Big(\hat{p}_t(\mathbf{s},\mathbf{h}) || q(\mathbf{s},\mathbf{h})\Big) \tag{3.16}$$

It can be verified that the above optimization problem can be replaced with,

$$q^{new}(\mathbf{h}) = \underset{q(\mathbf{h}) \in \ \mathcal{F}}{\arg\min}\, \mathrm{KL}\Big(\hat{p}_t(\mathbf{h}) || q(\mathbf{h})\Big) \tag{3.17}$$

and,

$$q^{new}(\mathbf{s}) = \underset{q(\mathbf{s})}{\arg\min}\, \mathrm{KL}\Big(\hat{p}_t(\mathbf{s}) || q(\mathbf{s})\Big) \tag{3.18}$$

where $\hat{p}_t(\mathbf{s})$ and $\hat{p}_t(\mathbf{h})$ are the respective marginal distributions of $\mathbf{s}$ and $\mathbf{h}$, derived from their joint distribution $\hat{p}_t(\mathbf{s},\mathbf{h})$. In the following we will solve (3.17) and (3.18) separately and use the solutions for updating the distributions of the corresponding factors.

Calculating $q^{new}(\mathbf{h})$ and updating $q_t(\mathbf{h})$

The marginal distribution of $\mathbf{h}$ is given by

$$\hat{p}_t(\mathbf{h}) = \sum_{\mathbf{s} \in \mathcal{A}_M^T} \hat{p}_t(\mathbf{s},\mathbf{h}) = \frac{1}{Z_t} q^{\setminus t}(\mathbf{h})\psi_t(\mathbf{h}) \tag{3.19}$$

where $\psi_t(\mathbf{h}) \triangleq \sum_{s_t \in \mathcal{A}_M} p(\mathbf{y}_t|s_t,\mathbf{h})$, and the normalization constant $Z_t = \mathbb{E}_{q^{\setminus t}(\mathbf{h})}[\psi_t(\mathbf{h})]$. Obviously, calculating Z_t by this formula will give us the same result as in (3.15). The intermediate distribution $\hat{p}_t(\mathbf{h})$ should be approximated by a pdf of the form $q(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\mathbf{m}_h, V_h)$

which satisfies (3.17). It turns out that the solution is obtained from the so-called *moment matching condition* [38], resulting in the new moments of $q(\mathbf{h})$ given by

$$\mathbf{m}_h^{new} = \mathbf{m}_h^{\setminus t} + V_h^{\setminus t} \nabla_m \tag{3.20}$$

and,

$$V_h^{new} = V_h^{\setminus t} - V_h^{\setminus t} (\nabla_m \nabla_m^T - 2\nabla_v) V_h^{\setminus t} \tag{3.21}$$

where, ∇_m and ∇_v are gradients defined as $\nabla_m \triangleq \left(\partial \log Z_t / \partial \mathbf{m}_h^{\setminus t}\right)^T$ and $\nabla_v \triangleq \left(\partial \log Z_t / \partial V_h^{\setminus t}\right)^T$, respectively. Using (3.15), it can be shown that (the proofs are omitted because of space limitation)

$$\nabla_m = \frac{1}{Z_t} \sum_{s_t \in \mathcal{A}_M} \left[\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big) S_t^T \Sigma_t^{-1} \boldsymbol{\zeta}_t \right] \tag{3.22}$$

and,

$$\begin{aligned} \nabla_v \;\; = \;\; & \frac{1}{2Z_t} \sum_{s_t \in \mathcal{A}_M} \Bigg[\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big) \times \\ & \Big(S_t^T \Sigma_t^{-1} \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t^T \Sigma_t^{-1} S_t - S_t^T \Sigma_t^{-1} S_t \Big) \Bigg] \end{aligned} \tag{3.23}$$

where, $\boldsymbol{\zeta}_t = \mathbf{y}_t - S_t \mathbf{m}_h^{\setminus t}$ and

$$\Sigma_t = S_t V_h^{\setminus t} S_t^T + \frac{1}{2} \sigma_w^2 I_{2N_r}. \tag{3.24}$$

After updating $q(\mathbf{h})$, its t-th factor $q_t(\mathbf{h})$, as in (3.8), is obtained from $q_t^{new}(\mathbf{h}) \propto q^{new}(\mathbf{h}) / q^{\setminus t}(\mathbf{h})$. Therefore, the new values for $\mathbf{m}_h(t)$ and $V_h(t)$ are given by

$$V_h^{new}(t) = \left(\left(V_h^{new}\right)^{-1} - \left(V_h^{\setminus t}\right)^{-1} \right)^{-1} \tag{3.25}$$

and,

$$\mathbf{m}_h^{new}(t) = V_h^{new}(t) \left(\left(V_h^{new}\right)^{-1} \mathbf{m}_h^{new} - \left(V_h^{\setminus t}\right)^{-1} \mathbf{m}_h^{\setminus t} \right). \tag{3.26}$$

Note that these updates do not depend on a specific value of the transmitted symbol s_t, and in fact a form of averaging over all possible values of this symbol appears in the formulas. □

Calculating $q^{new}(\mathbf{s})$ and updating $q_t(s_t)$

The marginal distribution of $\mathbf{s}$ is given by

$$\hat{p}_t(\mathbf{s}) = \int_{\mathbf{h}} \hat{p}_t(\mathbf{s},\mathbf{h})\mathrm{d}\mathbf{h} = \frac{1}{Z_t} q^{\setminus t}(\mathbf{s}) \mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big) \tag{3.27}$$

in which, Σ_t is given by (3.24). Therefore, the K-L divergence in (3.18) can be written as

$$\mathrm{KL}(\hat{p}_t(\mathbf{s})||q(\mathbf{s})) = \sum_{\mathbf{s}\in\mathcal{A}_M^T} \hat{p}_t(\mathbf{s}) \log \frac{\hat{p}_t(\mathbf{s})}{q(\mathbf{s})} \tag{3.28}$$

$$= \sum_{s_t \in \mathcal{A}_M} \frac{\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big)}{Z_t} \times \log \frac{\frac{\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big)}{Z_t}}{q_t(s_t)} \tag{3.29}$$

The minimizer of (3.29) for every $s_t \in \mathcal{A}_M$ and $t = 1, \ldots, T$ is given by

$$q_t^{new}(s_t) = \frac{1}{Z_t} \mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big) \tag{3.30}$$

Note that solving (3.18) only updates the pmf of the t-th symbol, namely $q_t(s_t)$. □

Each EP iteration involves updating all the T factors of $q_t(\mathbf{h})$ by (3.20) and (3.21) and $q_t^{new}(s_t)$ in (3.30). After the EP algorithm converges, $\mathbf{m}_h$ is adopted as the Maximum likelihood (ML) estimate of the unknown channel, i.e, $\mathbf{h}' = \mathbf{m}_h$. Consequently, the channel vector in (3.1) is given by

$$\tilde{\mathbf{h}}' = \mathbf{m}_{h,1:N_r} + j\mathbf{m}_{h,N_r+1:2N_r}. \tag{3.31}$$

Also, the following ML detection rule is employed for the transmitted symbols

$$s_t' = \underset{a\in\mathcal{A}_M}{\arg\max}\, q_t(a), \quad t = 1, \ldots, T. \tag{3.32}$$

As pointed out previously, the update rules for the moments of $q_t(\mathbf{h})$ in (3.25) and (3.26) reveal that they are independent of the specific value of the symbol s_t. In other words, refining the pdf factors for the channel is independent of refining the pdf factors for the transmitted symbols. This is to be expected since the channel vector is constant over all the received vectors[1]. Consequently, we do not need to updated the pmf of the symbols at each iteration. In fact, calculations of (3.30) for all symbols may be postponed to the end of the EP iterations, where the uncertainty about $\mathbf{h}$ is very small.

Now, at each iteration of the EP algorithm, the estimation of $\mathbf{h}$ gets closer to its actual value. This implies that the mean of $q(\mathbf{h})$ approaches the actual value of the channel vector $\mathbf{h}$ and the elements of the covariance matrix of $q(\mathbf{h})$ become smaller. Therefore, after an adequate number of iterations, the factors $q_t(\mathbf{h})$ become almost identical and independent of t. In other words, after a number of iterations, we will have $\mathbf{m}_h(1) \approx \ldots \approx \mathbf{m}_h(T)$ and $V_h(1) \approx \ldots \approx V_h(T)$. Inserting these values into (3.9) and (3.10), we get $\mathbf{m}_h = \mathbf{m}_h(1)$ and $V_h = V_h(1)/T$. Then, from (3.12) and (3.13) we can write, $\mathbf{m}_h^{\setminus t} = \mathbf{m}_h$ and $V_h^{\setminus t} = V_h(1)/(T-1) \approx V_h$. Therefore, (3.30) can be calculated after the final iteration of EP using these values of $\mathbf{m}_h^{\setminus t}$ and $V_h^{\setminus t}$. Moreover, since after an adequate number of iterations, the elements of $V_h^{\setminus t}$ will become very small, we can further simplify the calculation of $q_t(s_t)$ by replacing V_h with the all-zero matrix. Finally, for every $s_t \in \mathcal{A}_M$ and $t = 1, \ldots, T$, the pmf of the transmitted symbols can be calculated from

$$q_t(s_t) = \frac{\mathcal{N}\left(\mathbf{y}_t | S_t \mathbf{m}_h, \frac{1}{2}\sigma_w^2 I_{2N_r}\right)}{\sum_{s'_t \in \mathcal{A}_M} \mathcal{N}\left(\mathbf{y}_t | S'_t \mathbf{m}_h, \frac{1}{2}\sigma_w^2 I_{2N_r}\right)} \tag{3.33}$$

Therefore, the ML detection rule of (3.32) reduces to

$$s'_t = \arg\min_{s_t \in \mathcal{A}_M} ||\mathbf{y}_t - S_t \mathbf{m}_h||^2, \quad t = 1, \ldots, T. \tag{3.34}$$

[1]Note that this may not be true in the case of fast fading channels where the channel vector may change from symbol to symbol. In this case estimation of the channel coefficients at time t and detection of the transmitted symbol at time t will be interconnected.

Since at this stage the channel vector $\tilde{\mathbf{h}}'$ is already estimated and known, other alternative detection algorithms can also be used. For example, the receiver can employ the following MMSE detector

$$s'_t = \underset{s_t \in \mathcal{A}_M}{\arg\min} \left| s_t - \left(\tilde{\mathbf{h}}'^H \tilde{\mathbf{h}}' + \frac{\sigma_w^2}{E_s} \right)^{-1} \tilde{\mathbf{h}}' \tilde{\mathbf{y}}_t \right|, \quad t = 1, \ldots, T \tag{3.35}$$

The proposed procedure of channel estimation and symbol detection is summarized in Algorithm 2.

Data: The block of T received vectors and $p(\mathbf{y}_t|s_t, \mathbf{h})$ distributions in (3.4)
Result: A member of exponential family as (3.7) which is the closest pdf to (3.4), an estimation of the channel vector, and detected symbols
begin
  Initialize all $q_t(\mathbf{h})$ factors for $t = 1, \ldots, T$;
  Calculate $q(\mathbf{h})$ by (3.9) and (3.10);
  while *termination criteria has not been met* **do**
    for *t=1,...,T* **do**
      Calculate the cavity pdf by (3.12) and (3.13);
      Find $q^{new}(\mathbf{h})$ by (3.20) and (3.21);
      Update $q_t(\mathbf{h})$ by (3.25) and (3.26);
    end
  end
  Calculate the estimated channel vector by (3.31);
  Decode the symbols by ML rule in (3.34) or by MMSE rule in (3.35);
end

Algorithm 2: Noncoherent SIMO symbol detection by EP.

If the prior pdf $p(\mathbf{h})$ is available at the receiver, by considering $q'(\mathbf{h}) \propto p(\mathbf{h})q(\mathbf{h})$ from (3.6), the Bayesian estimation of the channel vector can be obtained as $\mathbf{h}'_B = (V_0^{-1} + V_h^{-1})^{-1}(V_0^{-1}\mathbf{m}_0 + V_h^{-1}\mathbf{m}_h)$. However, it is shown in the next section that the proposed noncoherent algorithm does not require any channel statistics such as the Rician K-factor.

The complexity of this algorithm is dominated by channel estimation part in equations (3.20), (3.21), (3.25), and (3.26), and is given by $\mathcal{O}(ITMN_r^3)$, where where I denotes the number of iterations of the EP algorithm. Considering the complexity of MMSE channel estimator as $\mathcal{O}(N_r^3)$ shows that the proposed EP algorithm is about ITM times more

complex than the pilot-based MMSE algorithm. However, as we will show in Section 3.4, the EP-based noncoherent algorithm can outperform MMSE estimator with only a single iteration ($I = 1$) over a fairly small block size T. Therefore, the overall complexity of the algorithm is not significantly higher.

3.4 Numerical Results

In this section we investigate the channel estimation and symbol detection performances of the proposed algorithm for the ML detector in (3.34) and the MMSE detector in (3.35), which will be referred to as EP-ML and EP-MMSE, respecticely. We compare the results with a coherent detector which first estimates the channel coefficients using the pilot-based MMSE algorithm and then uses that estimation for MMSE symbol detection. In all of our simulations, we have assumed a single pilot symbol for this coherent MMSE detector. To show the best possible achievable detection performance, we have also included the performance of the optimal ML receiver which has perfect knowledge of the CSI.

The differentially-encoded M-ary PSK modulation with symbol energy E_s is used in all the simulations. The signal-to-noise ratio (SNR) is defined as $10\log_{10}(E_s/\sigma_w^2)$[2]. For evaluating the channel estimation accuracy, the normalized estimation error between the channel vector $\mathbf{h}$ and its estimate $\mathbf{h}'$ is considered as $\delta_h = 10\log_{10}\left(||\mathbf{h}-\mathbf{h}'||^2/||\mathbf{h}||^2\right)$. Also, for measuring the accuracy in the estimation of the magnitude of the channel vector, we use $\delta_{|h|} = 10\log_{10}\left(\left|\left||\mathbf{h}|-|\mathbf{h}'|\right|\right|^2/\left|\left||\mathbf{h}|\right|\right|^2\right)$. For the EP algorithm, the mean vectors and covariance matrices of all the T factors of $\mathbf{h}$ are initialized as $m_h(t) = \mathbf{y}_1$ and $V_h(t) = I_{2N_r}$, respectively.

Fig. 3.1 shows the channel estimation performance of the proposed EP and the pilot-based MMSE estimators vs SNR for a SIMO system with $N_r = 100$ receiving antennas, which uses 8-DPSK modulation over Rayleigh fading channel ($K = 0$). The EP algorithm is run for a single iteration, i.e., I=1. As δ_h curves show, the pilot-based MMSE estimator outperforms the EP algorithm for all SNR values. This is mainly due to the inherent

[2]We assume that the channel power is one, i.e., $\mathbb{E}|h_i|^2 = 1$ for all i.

inability of noncoherent algorithms (such as the proposed EP detector) to remove the phase ambiguity in the channel. Also, due to this issue, increasing the block size results in only minor improvements in δ_h as can be seen in Fig. 3.1. Clearly for symbol detection, the effects of phase ambiguity can be alleviated by using a differential coding scheme. On the other hand, the $\delta_{|h|}$ curves show the advantage of the noncoherent EP estimator in estimating the magnitude of the channel vector. This figure shows that for $T = 2$, the performance of EP converges to that of MMSE as SNR increases. Moreover, increasing T from 2 to 20 and 50, results in estimation gains over the pilot-based MMSE of 10 dB and 15 dB, respectively. We would like to point out that the improved estimation of the magnitude of the channel is highly valuable in applications where power control is employed.

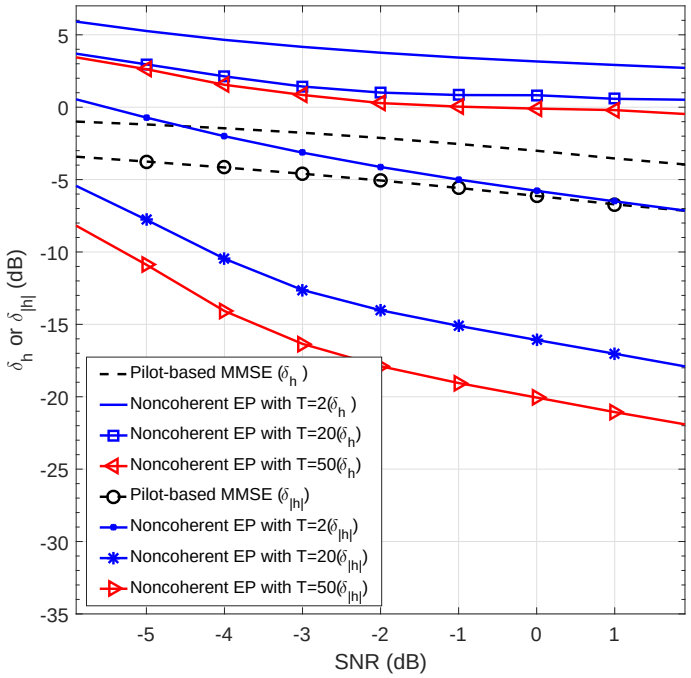


Figure 3.1: Channel estimation performance of a SIMO system with $N_r = 100$ and 8-DPSK, for pilot-based MMSE estimator, and the proposed EP estimator with $I = 1$ and block sizes $T = 2, 20, 50$.

The SER performance of the SIMO system of Fig. 3.1 for coherent ML, coherent MMSE, and noncoherent EP detectors is depicted in Fig. 3.2. This figure shows that the proposed EP detector outperforms the coherent MMSE detector. Moreover, for large values of $T = 20, 50$, the performance of noncoherent EP-ML detector is very close to the optimal

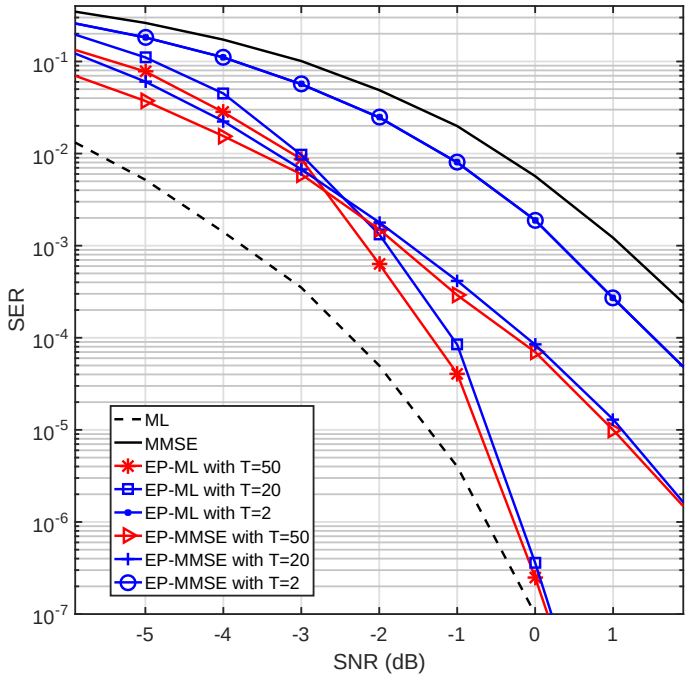


Figure 3.2: Detection performance of a SIMO system with $N_r = 100$ and 8-DPSK modulation, for coherent ML detector, coherent MMSE detector, and noncoherent EP detector with $I = 1$ and block sizes $T = 2, 20, 50$.

ML detector as SNR increases. Interestingly, this figure also shows that even for a block size of $T = 2$ symbols and with a single iteration, the proposed EP detector outperforms the pilot-based MMSE detector. It is important to note that the computational complexity of the detection method will be very small for such small values of T and I. In addition, the proposed method can also be used for fast fading channels with channel coherence times as low as two symbol durations.

Figs. 3.3 and 3.4 show the performance of channel estimation and symbol detection, respectively, for a SIMO system with $N_r = 100$ receiving antennas, $T = 5$ and 50 symbol blocks and 16-DPSK modulation over a Rayleigh fading channel. The number of iterations of the EP algorithm are $I = 1$ and $I = 4$. Fig. 3.3 shows that EP outperforms MMSE in the estimation of the magnitude of the channel coefficients. Moreover, increasing the number of iterations from 1 to 4 results in only a small improvement in the performance of EP. Fig. 3.4 shows that at $SER = 10^{-4}$, the noncoherent EP-ML/EP-MMSE outperforms the coherent MMSE by about 1 dB for $T = 5$ symbol blocks and EP-ML outperforms coherent

MMSE by about 3 dB for $T = 50$. This figure also shows that by increasing the size of the blocks to 50 symbols, SER performance of EP-ML converges to that of the optimal ML receiver. Comparing the SER curves of the two EP detectors reveals that increasing T results in only a small improvement in the performance of EP-MMSE. As before, increasing the number of iterations has a negligible effect on the SER performance.

Fig. 3.5 shows the SER performance of three SIMO configurations with $N_r = 10, 30,$ and 60 antennas which use the proposed EP detector over a Rician fading channel with a Rician factor of $K = 10$. Blocks of length $T = 5$ with 8-DPSK modulation are considered. Also all EP algorithms are assumed to work with $I = 1$ iteration and without any prior knowledge of K. This figure shows that for all configurations the proposed noncoherent algorithm outperforms the coherent MMSE detector.

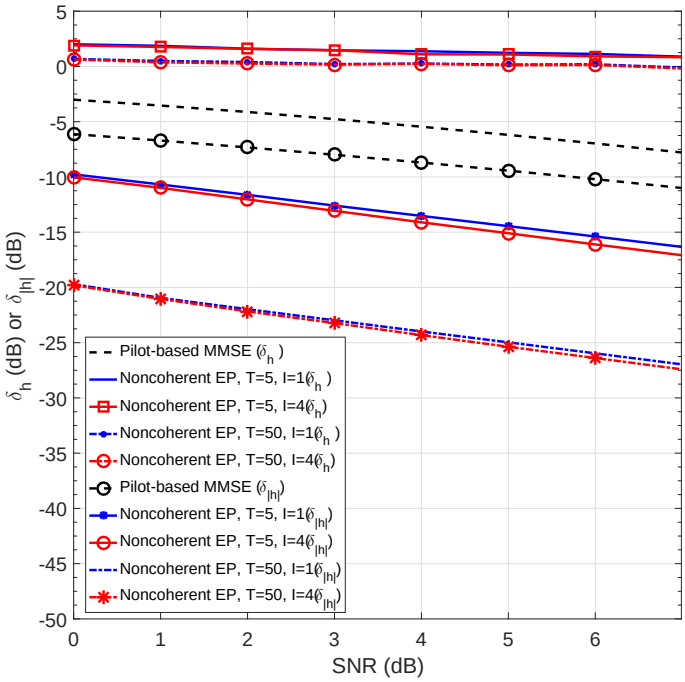


Figure 3.3: Channel estimation performance of a SIMO system with $N_r = 100$ and 16-DPSK modulation, for MMSE estimator with a single-symbol pilot, and EP estimators with block sizes $T = 5, 50$ and two different iterations $I = 1$ and $I = 4$.

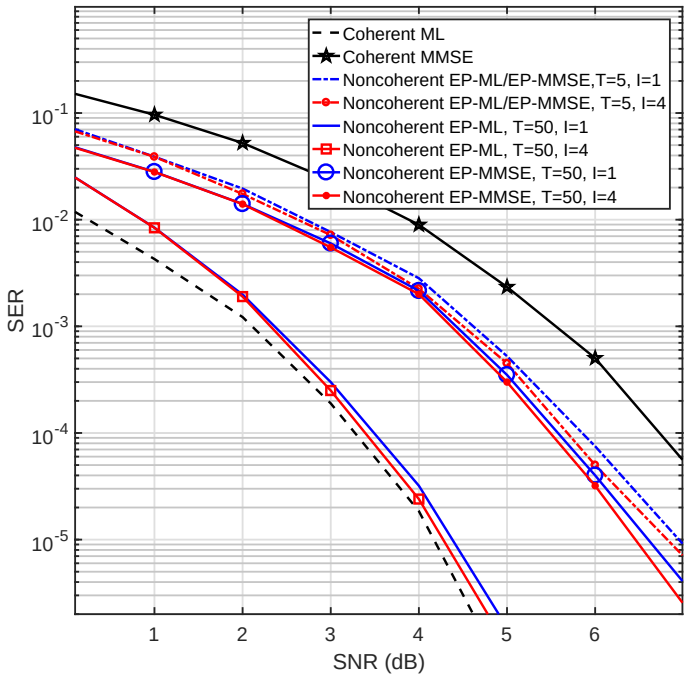


Figure 3.4: Detection performance of a SIMO system with $N_r = 100$ and 16-DPSK modulation, for MMSE estimator with a single-symbol pilot, and EP estimators with block sizes $T = 5, 50$ and two different iterations $I = 1$ and $I = 4$.

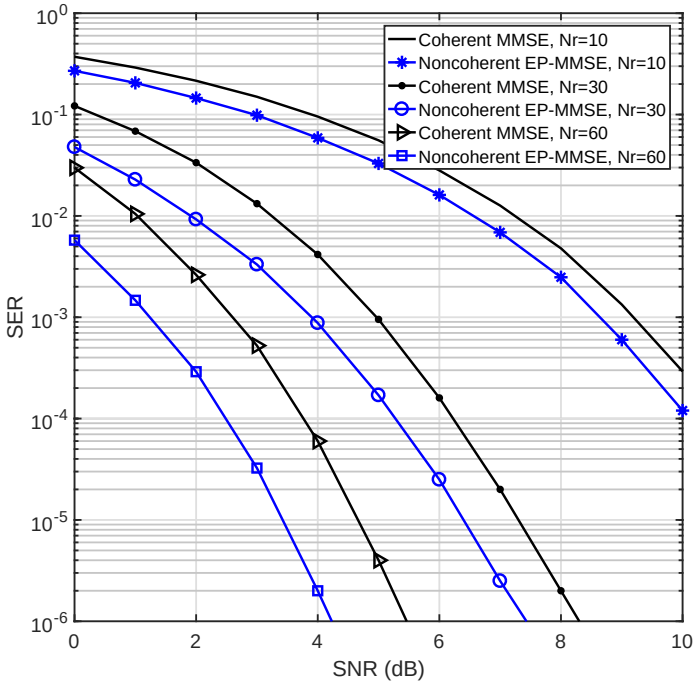


Figure 3.5: Detection performance of coherent MMSE and noncoherent EP detectors over Ricain fading channel with $K = 10$, 8-DPSK modulation, $I = 1$, $T = 5$, and three different antenna configurations.

Chapter 4
Joint Channel Estimation and Symbol Detection for Multi-Cell Massive MIMO Using Expectation Propagation

4.1 Introduction

Since the gains offered by multiple-input multiple-output (MIMO) systems scale with the number of transmitting and receiving antennas, research on high-order MIMO (also referred to as massive MIMO) systems has been accelerating in recent years [8, 9, 10, 11]. Massive MIMO is a multi-user MIMO (MU-MIMO) system in which the base station (BS) is equipped with an order-of-magnitude larger number of antennas compared to the traditional MIMO systems. Early studies have demonstrated the benefits of massive MIMO systems [12], and some field trials have been carried out to show the possibilities and limitations of this technology [13, 14, 15].

In order to realize the potential advantages of massive MIMO systems, several technical challenge must be addressed. Chief among them is the fact that symbol detection in the receiver requires accurate knowledge of channel state information (CSI) for a large number of channels [44]. Acquiring the CSI through transmission of pilot sequences, as is common in many wireless standards, is a resource-intensive process, due to the large number of channels involved. Moreover, since the number of transmitter antennas determines the minimum length of the pilot sequence, increasing the number of antennas in massive MIMO systems increases the length of pilot sequences. In particular, the required time for transmitting pilot symbols over the forward link may exceed the nominal coherence time of the channel. Consequently, deployment of massive MIMO systems in frequency division duplexing (FDD) mode remains an open problem [11]. Even in time division

duplexing (TDD) mode, due to lack of enough orthogonal sequences, pilots must be shared among the cells, giving rise to the so-called pilot contamination problem [9, 52]. This has motivated many researchers to investigate efficient noncoherent or pilotless symbol detection algorithms which do not require pilot transmission for CSI acquisition.

Noncoherent detection for single-input multiple-output (SIMO) systems is suggested in several previous works. For instance, [45] and [46] propose techniques based on sphere decoding with an efficient search radius. Another noncoherent technique based on a proposed modulation technique in which information is carried in the power of the transmitted symbols over uncorrelated Rician fading channel, is suggested in [47] and [48]. In [49] the authors propose an optimal constellation for achieving the minimum noncoherent detection error, where the dependency on large scale channel statistics is reduced by applying proper constellation design. Another algorithm in [50] proposes a noncoherent scheme for large scale SIMO systems based on the knowledge of the average received powers from different users. In [53] we have developed a joint channel estimation and symbol detection algorithm for SIMO systems based on Expectation Propagation (EP). It is shown in this work that the proposed detector outperforms the pilot-based MMSE detector for both Rayleigh and Rician fading channels.

When the ratio of the number of transmitting antennas to the number of receiving antennas is small, as is common in the forward link of massive MIMO systems, there will be a large degree of freedom which can be exploited by noncoherent algorithms. For example, [54] suggests a noncoherent channel estimation technique for multi-cell massive MIMO systems based on the eigenvalue decomposition (EVD) of the correlation matrix of the received vectors. However, the proposed algorithm is sensitive to the size of the antenna array as well as the accuracy of the empirically calculated correlation matrix of the received vectors. A noncoherent algorithm based on subspace projection and random matrix theory is suggested in [55, 56, 57]. Assuming that the number of users per cell remains fixed, the authors show that under certain conditions on the powers of the trans-

mitting users, the spectrum of the sample covariance matrix asymptotically decomposes into the signal eigenvalue spectrum and the interference-plus-noise eigenvalue spectrum as the number of BS antennas grows [56]. The limiting support of the two spectra are approximately characterized and a bound on the power difference of the signal and interference is determined in order for the two to become disjoint. It is noted in [58] that this bound is independent of the noise variance and, as a result, becomes inaccurate under low signal-to-noise ratio (SNR) regime. A new asymptotic condition on spectrum separability is then derived in [58] by considering the exact asymptotic characterization of interference-plus-noise spectrum. These algorithms reduce pilot contamination by appropriate power control and estimate the channel matrix of the desired cell. Finally [59] exploits the sparsity of the massive MIMO channels and transforms the channel estimation problem into learning on a Gaussian mixture model, which can be solved using algorithms such as Expectation Maximization.

Coherent MIMO detection based on EP was recently suggested in [32], where the performance of the algorithm is evaluated on the premise that perfect channel state information (CSI) is available at the receiver. However, as mentioned earlier, providing perfect CSI in massive MIMO systems is a challenging problem. At best, only a noisy estimate of the channel coefficients will be available at the receiver. Also, other detrimental effects, such as pilot contamination, aging and quantization errors, limit the accuracy of the CSI estimates. In [60] we show that although channel estimation improves by increasing the signal-to-noise ratio (SNR)[1], surprisingly, at high SNR values, the rate of improvement of symbol error rate (SER) vs. SNR decreases. We investigate this behavior of the EP detector in the case of imperfect CSI and propose a modified detector in order to recover some of the performance loss of the EP detector. Simulation results verify that the proposed modification improves the performance of EP in the case of imperfect CSI, particularly in higher SNR regions, and that for the modified detector, the slope of the SER vs. SNR plots are similar

[1]The SNR of the pilot and information sequences are assumed to be the same

to the case of perfect CSI.

In this work we propose a noncoherent channel estimation and symbol detection technique for multi-cell massive MIMO systems based on the Expectation Propagation (EP) algorithm [38, 31]. Since the channel matrix and the transmitted vectors are both unknown, the proposed EP-based algorithm is applied on a hybrid model [51], and iteratively searches for the best approximation of the joint probability density function (PDF) of these unknowns. The output PDF is used for direct estimation of the channel coefficients, as well as the transmitted symbols. To overcome the inherent ambiguity of noncoherent estimators, we use the output of the EVD-based estimator in [54] to initialize the channel coefficients in our algorithm. This allows for the EP-based receiver to improve its performance over the EVD-based algorithm. We show that the proposed EP estimator outperforms the EVD-based algorithm in both channel estimation and symbol detection. Moreover, this approach diminishes the sensitivity of the EVD-based algorithm to the array and transmission block sizes.

The rest of this chapter is organized as follows. The system model is presented in Section 4.2. Section 4.3 covers a brief review of the EVD-based algorithm. The proposed algorithm is presented in Section 4.4. Finally, simulation results are presented in Section 4.5.

Notations: Throughout this paper, small letters (x) are used for scalars, bold small letters ($\mathbf{x}$) for vectors, and capital letters (X) for matrices. $\mathbb{R}$ and $\mathbb{C}$ represent the set of real and complex numbers, respectively. $\mathfrak{R}(z)$ and $\mathfrak{I}(z)$ denote the real and imaginary parts of the complex variable z. For a set of complex variables $A = \{z_1, z_2, \cdots\}$, we denote $\mathfrak{R}(A) \triangleq \{\mathfrak{R}(z_1), \mathfrak{R}(z_2), \cdots\}$ and $\mathfrak{I}(A) \triangleq \{\mathfrak{I}(z_1), \mathfrak{I}(z_2), \cdots\}$. The superscripts $(.)^T$, $(.)^H$, and $(.)^{-1}$ represent transpose, Hermitian transpose, and matrix inverse, respectively. Also, $\otimes$ and $\circ$ denote the matrix Kronecker and Hadamard products, respectively. For a probability density function (PDF) $p(.)$, $\mathbb{E}_p$ denotes the expectation operator with respect to $p(.)$. I_N denotes the $N \times N$ identity matrix and $\text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$ denotes the $n \times n$

diagonal matrix with $\lambda_1, \lambda_2, \ldots, \lambda_n$ as its main diagonal. Finally, vec(A), $\|\mathbf{a}\|$, and $\|A\|_F$ denote the vectorization of the matrix A, the ℓ^2 norm of vector $\mathbf{a}$, and the Frobenius norm of matrix A, respectively.

4.2 System Model

Consider a multi-user MIMO network consisting of L cells each served with its own BS and with K users in each cell. Assume the base stations have M antennas and all users have single-antenna transceivers. A simple schematic of this network is demonstrated in Fig. 4.1. The channel gain between the m-th antenna of the l-th BS and the k-th user located in the i-th cell is denoted by $\tilde{h}_{limk}$. Each channel coefficient $\tilde{h}_{limk}$ can be decomposed into two parts as

$$\tilde{h}_{limk} = \tilde{g}_{limk}\sqrt{\beta_{lik}}, \tag{4.1}$$

in which, $\tilde{g}_{limk}$ denotes the fast fading coefficient from the k-th user in cell i to the m-th antenna of BS l, and β_{lik} represents the geometric attenuation and the shadowing effects which is assumed to be independent of the antenna index m and to be constant and known a priori.

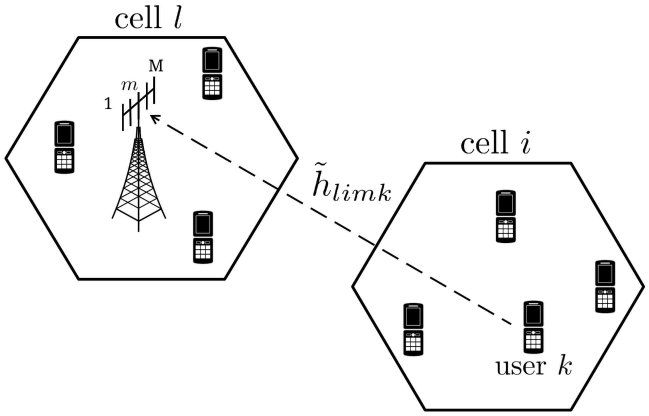


Figure 4.1: Multi-cell multi-user MIMO network.

The $M \times 1$ fast fading vector from user k in cell i to the BS antenna array at cell l is denoted by $\tilde{\mathbf{g}}_{lik} = [\tilde{g}_{li1k}, \tilde{g}_{li2k}, \ldots, \tilde{g}_{liMk}]^T$, and the $M \times K$ fading matrix from all the users of cell i to the l-th BS is denoted by $\tilde{G}_{li} = [\tilde{\mathbf{g}}_{li1}, \ldots, \tilde{\mathbf{g}}_{liK}]$. Consequently, the total channel

gain between users in cell i and the BS in cell l will be an $M \times K$ matrix denoted by $\tilde{H}_{li}$ and given by

$$\tilde{H}_{li} \triangleq \tilde{G}_{li} D_{li}^{\frac{1}{2}} \tag{4.2}$$

where $D_{li} \triangleq \text{diag}[\beta_{li1}, \beta_{li2}, \ldots, \beta_{liK}]$.

Assuming that the fast fading coefficients between different channels and/or users are independent and identically distributed (iid) with $\mathbb{E}[\tilde{g}_{limk}\tilde{g}^*_{l'i'm'k'}] = \delta_{ll'}\delta_{ii'}\delta_{mm'}\delta_{kk'}$, we get $\mathbb{E}[\tilde{G}_{li}\tilde{G}_{lj}^H] = KI_M\delta_{ij}$. Therefore,

$$\mathbb{E}[\tilde{H}_{li}\tilde{H}_{lj}^H] = \mathbb{E}[\tilde{H}_{li}\tilde{H}_{li}^H]\delta_{ij} = \mathbb{E}[\tilde{G}_{li}D_{li}\tilde{G}_{li}^H]\delta_{ij} \tag{4.3}$$

$$= \sum_{k=1}^{K} \beta_{lik}\mathbb{E}[\tilde{\mathbf{g}}_{lik}\tilde{\mathbf{g}}_{lik}^H]\delta_{ij} = \Big(\sum_{k=1}^{K}\beta_{lik}\Big) I_M\delta_{ij} \tag{4.4}$$

Without loss of generality, suppose we are interested in the first cell, i.e., $l = 1$. Therefore, the received vector at this BS at t-th channel use is given by

$$\tilde{\mathbf{y}}_1(t) = \sum_{i=1}^{L} \tilde{H}_{1i}\tilde{\mathbf{s}}_i(t) + \tilde{\mathbf{w}}_1(t) \tag{4.5}$$

$$= \tilde{H}_{11}\tilde{\mathbf{s}}_1(t) + \sum_{i=2}^{L} \tilde{H}_{1i}\tilde{\mathbf{s}}_i(t) + \tilde{\mathbf{w}}_1(t) \tag{4.6}$$

where $\tilde{\mathbf{s}}_i(t) = [\tilde{s}_{i1}(t), \tilde{s}_{i2}(t), \ldots, \tilde{s}_{iK}(t)]^T$ is the vector of transmitted symbols by users in cell i. The symbols $\tilde{s}_{ij}(t)$ are assumed to be independently selected from an $\mathcal{M}$-ary modulation constellation denoted as $\tilde{\mathcal{A}}_{\mathcal{M}}$, with zero mean and average energy E_s. Therefore we have $\mathbb{E}\big[\tilde{\mathbf{s}}_i(t)\tilde{\mathbf{s}}_j(t)^H\big] = E_s\delta_{ij}I_K$. Also, $\tilde{\mathbf{w}}_l(t) \sim \mathcal{CN}(\tilde{\mathbf{w}}_l|\mathbf{0}, I_M)$ is the circularly symmetric additive white Gaussian noise at the l-th BS and are assumed to be independent for different BS's. Setting $\tilde{\mathbf{w}}'_1(t) \triangleq \sum_{i=2}^{L} \tilde{H}_{1i}\tilde{\mathbf{s}}_i(t) + \tilde{\mathbf{w}}_1(t)$ as the combination of interference plus noise at BS 1, we can write

$$\tilde{\mathbf{y}}_1(t) = \tilde{H}_{11}\tilde{\mathbf{s}}_1(t) + \tilde{\mathbf{w}}'_1(t) \tag{4.7}$$

We can treat $\tilde{\mathbf{w}}_1'(t)$ as a new additive noise vector. The mean of $\tilde{\mathbf{w}}_1'(t)$ is given by $\mathbb{E}[\tilde{\mathbf{w}}_1'(t)] = \mathbf{0}$ and its covariance matrix can be calculated as follows.

$$R_w' \triangleq \mathbb{E}[\tilde{\mathbf{w}}_1'(t)\big(\tilde{\mathbf{w}}_1'(t)\big)^H] \tag{4.8}$$

$$= \sum_{i=2}^{L}\sum_{j=2}^{L} \mathbb{E}\left[\tilde{H}_{1i}\mathbb{E}\left[\tilde{\mathbf{s}}_i(t)\tilde{\mathbf{s}}_j(t)^H\right]\tilde{H}_{1j}^H\right] + I_M \tag{4.9}$$

$$= E_s \sum_{i=2}^{L} \mathbb{E}\Big[\tilde{H}_{1i}\tilde{H}_{1i}^H\Big] + I_M \tag{4.10}$$

$$= E_s \sum_{i=2}^{L} \Big(\sum_{k=1}^{K} \beta_{1ik} I_M\Big) + I_M \tag{4.11}$$

$$= \Big(E_s \sum_{i=2}^{L}\sum_{k=1}^{K} \beta_{1ik} + 1\Big) I_M \tag{4.12}$$

Assuming that the product KL is large, we invoke the central limit theorem and assume that the mulit-user interference $\sum_{i=2}^{L} \tilde{H}_{1i}\tilde{\mathbf{s}}_i(t)$ is a Gaussian random vector. It then follows that the received vector at the first BS is given by (4.7), where $\tilde{\mathbf{w}}_1'(t) \sim \mathcal{CN}\Big(\tilde{\mathbf{w}}_1'|\mathbf{0}, \sigma_w'^2 I_M\Big)$, and where

$$\sigma_w'^2 \triangleq E_s \sum_{i=2}^{L}\sum_{k=1}^{K} \beta_{1ik} + 1 \tag{4.13}$$

To simplify our notations, for the received vector of the target BS in a multi-cell multi-user model at t-th channel use, in the sequel we rewrite (4.7) as

$$\tilde{\mathbf{y}}_t = \tilde{H}\tilde{\mathbf{s}}_t + \tilde{\mathbf{w}}_t', \tag{4.14}$$

in which, $\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K$ is the vector of transmitted symbols with entries in $\tilde{\mathcal{A}}_{\mathcal{M}}$, $\tilde{H} \in \mathbb{C}^{M\times K}$ is the channel matrix, and $\tilde{\mathbf{w}}_t' \in \mathbb{C}^{M\times 1}$ is the additive noise vector such that $\tilde{\mathbf{w}}_t' \sim \mathcal{CN}(\tilde{\mathbf{w}}_t'|\mathbf{0}, \sigma_w'^2 I_M)$ with $\sigma_w'^2$ given in (4.13).

By applying the vectorization property $\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B)$, (4.14) can be expressed as

$$\tilde{\mathbf{y}}_t = \tilde{S}_t\tilde{\mathbf{h}} + \tilde{\mathbf{w}}_t' \tag{4.15}$$

in which, $\tilde{S}_t \triangleq \tilde{\mathbf{s}}_t^T \otimes I_M$ and $\tilde{\mathbf{h}} \triangleq \text{vec}(\tilde{H})$. The all-real equivalent of (4.15) is given by

$$\begin{bmatrix} \Re(\tilde{\mathbf{y}}_t) \\ \Im(\tilde{\mathbf{y}}_t) \end{bmatrix} = \begin{bmatrix} \Re(\tilde{S}_t) & -\Im(\tilde{S}_t) \\ \Im(\tilde{S}_t) & \Re(\tilde{S}_t) \end{bmatrix} \begin{bmatrix} \Re(\tilde{\mathbf{h}}) \\ \Im(\tilde{\mathbf{h}}) \end{bmatrix} + \begin{bmatrix} \Re(\tilde{\mathbf{w}}'_t) \\ \Im(\tilde{\mathbf{w}}'_t) \end{bmatrix} \tag{4.16}$$

Denoting $\mathbf{y}_t \triangleq \left(\Re(\tilde{\mathbf{y}}_t^T)\ \Im(\tilde{\mathbf{y}}_t^T)\right)^T$, $S_t \triangleq \begin{bmatrix} \Re(\tilde{S}_t) & -\Im(\tilde{S}_t) \\ \Im(\tilde{S}_t) & \Re(\tilde{S}_t) \end{bmatrix}$, $\mathbf{h} \triangleq \left(\Re(\tilde{\mathbf{h}}^T)\ \Im(\tilde{\mathbf{h}}^T)\right)^T$, and $\mathbf{w}'_t \triangleq \left(\Re(\tilde{\mathbf{w}}_t'^T)\ \Im(\tilde{\mathbf{w}}_t'^T)\right)^T$, (4.16) can be written as

$$\mathbf{y}_t = S_t\mathbf{h} + \mathbf{w}'_t. \tag{4.17}$$

Finally it follows from the properties of circularly symmetric Gaussian random vectors that $\mathbf{w}'_t \sim \mathcal{N}(\mathbf{w}'|\mathbf{0}, \frac{1}{2}\sigma_w'^2 I_{2M})$.

For the noncoherent detector being considered here, the channel vector, $\mathbf{h}$, as well as the transmitted symbols, S_t, are unknown. Therefore, a joint symbol detection and channel estimation algorithm must be employed at the receiver to estimate the channel vector and detect the transmitted symbols.

Consider a block of T transmitted vectors $\tilde{S} \triangleq [\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_T]$ and the corresponding received vectors $Y \triangleq [\mathbf{y}_1\ \ldots\ \mathbf{y}_T]$. We assume a block fading channel where the channel matrices remain unchanged for the duration of a transmission block consisting of T transmitted symbols. The posterior distribution of $\tilde{S}$ and $\mathbf{h}$ is given by

$$\begin{aligned} p(\tilde{S}, \mathbf{h}|Y) &\propto p(\tilde{S}, \mathbf{h})p(Y|\tilde{S}, \mathbf{h}) \\ &\propto \left(p(\mathbf{h})\prod_{t=1}^{T} p(\tilde{\mathbf{s}}_t)\right)\prod_{t=1}^{T} p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h}) \end{aligned} \tag{4.18}$$

in which $p(\tilde{\mathbf{s}}_t)$ is the probability mass function (pmf) of the t-th transmitted vector, $p(\mathbf{h})$ is the prior PDF of the channel vector,[2] $p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h})$ is the a posteriori distribution of the

[2]In developing the noncoherent receiver, we assume that $\mathbf{h}$ is an unknown constant vector during the

t-th received vector, which from (4.17) is given by $p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h}) = \mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{h}, \frac{1}{2}\sigma_w'^2 I_{2M}\Big)$.[3]

The optimum receiver finds the unknowns as the maximizers of (4.18), namely,

$$(\tilde{S}, \mathbf{h})^* = \underset{\tilde{S}\in\tilde{\mathcal{A}}_{\mathcal{M}}^{K\times T}, \mathbf{h}\in\mathbb{R}^{2M}}{\arg\max} \; p(\tilde{S}, \mathbf{h}|Y) \tag{4.19}$$

The optimization in (4.19) is not tractable due to the complexity of the posterior distribution in (4.18). In fact the optimization requires the computation of marginal distributions which demand extensive multi-dimensional numerical integrations. In Section 4.4 we develop an estimation technique based on the EP algorithm. This approach exploits the multiplicative nature of (4.18) to find an approximation for the posterior distribution of $(\tilde{S}, \mathbf{h})$ so that the marginals can be easily computed.

4.3 Review of the EVD-Based Massive MIMO Channel Estimation

A channel estimation algorithm based on the existing degrees of freedom in massive MIMO systems is suggested in [54]. The proposed algorithm estimates the channel vectors from the eigenvectors of the sample correlation matrix of the received vectors at the BS.

Suppose that we are interested in estimating the channel matrix of the first cell in the multi-user multi-cell network. The received vector at the first BS at time t is given by (4.5). The correlation matrix of this vector is given by

$$\begin{aligned} R_y &\triangleq \mathbb{E}[\tilde{\mathbf{y}}_1(t)\tilde{\mathbf{y}}_1^H(t)] \\ &= \mathbb{E}\big[\big(\sum_{i=1}^{L}\tilde{H}_{1i}\tilde{\mathbf{s}}_i(t) + \tilde{\mathbf{w}}_1(t)\big)\big(\sum_{j=1}^{L}\tilde{H}_{1j}\tilde{\mathbf{s}}_j(t) + \tilde{\mathbf{w}}_1(t)\big)^H\big] \\ &= E_s\sum_{i=1}^{L}\tilde{H}_{1i}\tilde{H}_{1i}^H + I_M = E_s\sum_{i=1}^{L}\tilde{G}_{1i}D_{1i}\tilde{G}_{1i}^H + I_M \end{aligned} \tag{4.20}$$

observation period T. However, to estimate it using the EP algorithm, we estimate a PDF for $\mathbf{h}$ from which the ML estimate of $\mathbf{h}$ is obtained.

[3]Note that the matrix S_t only depends on the vector $\tilde{\mathbf{s}}_t$. Therefore, for the conditional PDF of $\mathbf{y}_t$ we can write $p(\mathbf{y}_t|S_t, \mathbf{h}) = p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h})$.

Under the so-called *favorable propagation condition*, [9], the channel vectors of different terminals will be mutually orthogonal, i.e., as $M \longrightarrow \infty$,

$$\frac{1}{M}\tilde{G}_{li}^{H}\tilde{G}_{lj} \longrightarrow I_K \delta_{ij} \tag{4.21}$$

Therefore, multiplying R_y by $\tilde{G}_{11}$ and using (4.21) we get

$$\begin{aligned} R_y\tilde{G}_{11} &\approx E_s \sum_{i=1}^{L} \tilde{G}_{1i} D_{1i} \tilde{G}_{1i}^{H} \tilde{G}_{11} + \tilde{G}_{11} \\ &= \tilde{G}_{11}(E_s M D_{11} + I_K). \end{aligned} \tag{4.22}$$

Note that $\Lambda_{11} \triangleq E_s M D_{11} + I_K$ is a diagonal matrix with $\{E_s M\beta_{11k} + 1,\ k = 1, \ldots, K\}$ as its main diagonal. Therefore, $R_y\tilde{G}_{11} = \tilde{G}_{11}\Lambda_{11}$ along with the pairwise orthogonality of columns of $\tilde{G}_{11}$ shows that (4.22) can be considered as the characteristic equation of the correlation matrix R_y. Therefore, the kth column of $\tilde{G}_{11}$ is proportional to the kth eigenvector of R_y corresponding to the eigenvalue $E_s M\beta_{11k}+1$. Assuming that $E_s M\beta_{11k}+1$, $k = 1, 2, \ldots, K$ are distinct and known a priori, the ordering of the eigenvectors among the K users can be determined up to a constant factor. In other words, if U_1 is the $M \times K$ matrix whose columns are the eigenvectors of R_y, then the estimate of $\tilde{G}_{11}$, denoted by $\hat{G}_{11}$ is given by $\hat{G}_{11} = U_1 C_1$, where C_1 is a diagonal matrix. Since C_1 is unknown at the BS, it is estimated using a short pilot sequence.

Suppose each user transmits τ pilot symbols. Assuming that the pilots of different cells are mutually orthogonal, the received $M \times \tau$ matrix at the 1st BS is given by

$$\tilde{Y}_1 = \tilde{H}_{11} P_1 + \tilde{W}_1 \tag{4.23}$$

where P_1 is the $K \times \tau$ matrix of pilots from the K users in cell 1, and $\tilde{W}_1$ is the noise matrix. The ambiguity matrix C_1 can now be estimated by solving the following optimization

problem

$$\hat{C}_1 = \arg\min_{C_1 \in \Theta} \|\tilde{Y}_1 - U_1 C_1 D_{11}^{1/2} P_1\|_F^2, \tag{4.24}$$

in which Θ is the set of all $K \times K$ diagonal matrices in $\mathbb{C}^{K \times K}$.

Denoting $J \triangleq \|\tilde{Y}_1 - U_1 C_1 D_{11}^{1/2} P_1\|\|_F^2$ and considering the optimization constraint, the derivative of J with respect to C_1 is given by

$$\frac{\partial J}{\partial C_1} = (D_{11}^{1/2} P_1 P_1^H D_{11}^{1/2} C_1^H U_1^H U_1 - D_{11}^{1/2} P_1 \tilde{Y}_1^H U_1) \circ I_K \tag{4.25}$$

Solving $\partial J / \partial C_1 = \mathbf{0}$ leads to

$$(\Xi \hat{C}_1^H \Phi) \circ I_K = \Psi \circ I_K, \tag{4.26}$$

in which, $\Xi = D_{11}^{1/2} P_1 P_1^H D_{11}^{1/2}$, $\Phi = U_1^H U_1$, and $\Psi = D_{11}^{1/2} P_1 \tilde{Y}_1^H U_1$. Now, suppose the vector of K diagonal elements of $\hat{C}_1$ is denoted by $\hat{\mathbf{c}}_1$. Then, it can be easily shown that the solution of (4.26) is

$$\hat{\mathbf{c}}_1 = (\Phi^H \circ \Xi^*)^{-1} (\Psi \circ I_K) \mathbf{1}_K, \tag{4.27}$$

in which, $\mathbf{1}_K$ is a $K \times 1$ all-one vector.

After calculating the constant matrix, the channel can be estimated as

$$\hat{H}_{11} = U_1 \hat{C}_1 D_{11}^{\frac{1}{2}} \tag{4.28}$$

In practice the correlation matrix R_y is not available and the sample covariance matrix is used instead. Using a block of N samples, the sample correlation matrix can be calculated as

$$\hat{R}_y = \frac{1}{N} \sum_{n=1}^{N} \tilde{\mathbf{y}}_1(n) \tilde{\mathbf{y}}_1^H(n) \tag{4.29}$$

As mentioned in [54], error in the calculated sample correlation matrix due to the

limited block size N and lack of true channel mutual orthogonality, particularly for smaller antenna arrays, are two major sources of error in the proposed algorithm. In this paper we develop an EP-based noncoherent detector which is initialized with the output of the EVD-based detector. After a few iterations, the proposed detector significantly improves the performance of the EVD-based detector.

4.4 EP Formulation for Noncoherent Detection

Let $\mathcal{F}$ denote a family of exponential distributions. Using EP we exploit the factorized structure of (4.18) to approximate the posterior distribution $p(\tilde{S}, \mathbf{h}|Y)$ with distributions from $\mathcal{F}$ (for a review of the EP algorithm please refer to [38, 31]). To this end, we employ the following approximation for the posterior distribution $p(\tilde{S}, \mathbf{h}|Y)$,

$$q'(\tilde{S}, \mathbf{h}) = \left(p(\mathbf{h}) \prod_{t=1}^{T} p(\tilde{\mathbf{s}}_t) \right) q(\tilde{S}, \mathbf{h}) \tag{4.30}$$

in which $q(\tilde{S}, \mathbf{h}) \in \mathcal{F}$ is considered as an approximation to the likelihood function $p(Y|\tilde{S}, \mathbf{h})$. Since the channel vector is a continuous vector and the transmitted symbols are discrete random variables, this will be a hybrid model [51]. After calculating $q(\tilde{S}, \mathbf{h})$ from EP as the best approximation (in $\mathcal{F}$) to $p(Y|\tilde{S}, \mathbf{h})$, an approximation to the a posterior distribution can be calculated using (4.30). However, assuming independent, identically distributed (iid) transmitted vectors, the solution to (4.19) can be equivalently obtained from maximizing the likelihood function $p(Y|\tilde{S}, \mathbf{h})$. Therefore, we do not need to evaluate (4.30). The transmitted vectors and channel coefficients can be estimated by maximizing $q(\tilde{S}, \mathbf{h})$.

The proposed approximation $q(\tilde{S}, h)$ is further written in a factorized form as follows.

$$q(\tilde{S}, \mathbf{h}) \propto \prod_{t=1}^{T} q_t(\tilde{\mathbf{s}}_t, \mathbf{h}) \propto \prod_{t=1}^{T} q_t(\tilde{\mathbf{s}}_t) q_t(\mathbf{h}) \tag{4.31}$$

$$\propto q(\tilde{S}) q(\mathbf{h}) \tag{4.32}$$

in which, $q(\mathbf{h}) \propto \prod_{t=1}^{T} q_t(\mathbf{h})$ and $q(\tilde{S}) = \prod_{t=1}^{T} q_t(\tilde{\mathbf{s}}_t)$. We note that by selecting each factor

$q_t(\mathbf{h})$ from $\mathcal{F}$, we are ensured that their product, $q(\mathbf{h})$, is also in $\mathcal{F}$. Moreover, multiplying $q(\mathbf{h})$ by the pmf $q(\tilde{S})$ will not change its structure. Therefore $q(\tilde{S}, \mathbf{h})$ will also be in the family of exponential distributions.

We consider the family of multivariate Gaussian distributions for $\mathcal{F}$. In particular, we assume that all $q_t(\mathbf{h})$ factors are multivariate Gaussian with mean vector $\mathbf{m}_h(t)$ and covariance matrix $V_h(t)$, i.e., $q_t(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\mathbf{m}_h(t), V_h(t))$. Therefore their product will be proportional to a multivariate Gaussian distribution $q(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\mathbf{m}_h, V_h)$, with covariance matrix

$$V_h = \left(\sum_{t=1}^{T} V_h(t)^{-1} \right)^{-1} \tag{4.33}$$

and mean vector

$$\mathbf{m}_h = V_h \sum_{t=1}^{T} V_h(t)^{-1} \mathbf{m}_h(t). \tag{4.34}$$

Also, all $q(\tilde{\mathbf{s}}_t)$ factors are assumed to be pmf of their corresponding random vectors. Since $\tilde{\mathbf{s}}_t$ is a K-dimensional vector with elements from $\tilde{\mathcal{A}}_{\mathcal{M}}$, for each possible vector $\tilde{\mathbf{s}}_t$, $q(\tilde{\mathbf{s}}_t)$ will be a set of $\mathcal{M}^K$ probabilities. Denoting the set of all K-tuples over $\tilde{\mathcal{A}}_{\mathcal{M}}$ by $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_{\mathcal{M}^K}\}$, then $q(\tilde{\mathbf{s}}_t)$ can be defined as follows

$$q(\tilde{\mathbf{s}}_t) = \Big(\mathbb{P}[\tilde{\mathbf{s}}_t = \mathbf{a}_1], \ldots, \mathbb{P}[\tilde{\mathbf{s}}_t = \mathbf{a}_{\mathcal{M}^K}] \Big). \tag{4.35}$$

EP uses the t-th factor in the likelihood function $p(Y|\tilde{S}, \mathbf{h})$, i.e., $p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h})$, for refining the t-th approximating distribution $q_t(\tilde{\mathbf{s}}_t, \mathbf{h})$. Toward this, the so called t-th cavity distribution is calculated as

$$q^{\setminus t}(\tilde{S}, \mathbf{h}) \propto \frac{q(\tilde{S}, \mathbf{h})}{q_t(\tilde{\mathbf{s}}_t, \mathbf{h})} = \frac{q(\mathbf{h}) q(\tilde{S})}{q_t(\mathbf{h}) q_t(\tilde{\mathbf{s}}_t)} \tag{4.36}$$

$$\propto q^{\setminus t}(\mathbf{h}) q^{\setminus t}(\tilde{S}) \tag{4.37}$$

in which,

$$q^{\setminus t}(\mathbf{h}) \propto \frac{q(\mathbf{h})}{q_t(\mathbf{h})} = \frac{\mathcal{N}(\mathbf{h}|\mathbf{m}_h, V_h)}{\mathcal{N}(\mathbf{h}|\mathbf{m}_h(t), V_h(t))} \tag{4.38}$$

$$\propto \mathcal{N}(\mathbf{h}|\mathbf{m}_h^{\setminus t}, V_h^{\setminus t}) \tag{4.39}$$

where it can be shown that

$$V_h^{\setminus t} = (V_h^{-1} - V_h(t)^{-1})^{-1} \tag{4.40}$$

$$\mathbf{m}_h^{\setminus t} = V_h^{\setminus t}\Big(V_h^{-1}\mathbf{m}_h - V_h(t)^{-1}\mathbf{m}_h(t)\Big). \tag{4.41}$$

Moreover,

$$q^{\setminus t}(\tilde{S}) \propto \frac{q(\tilde{S})}{q_t(\tilde{\mathbf{s}}_t)} = \frac{\prod_{i=1}^{T} q_i(\tilde{\mathbf{s}}_i)}{q_t(\tilde{\mathbf{s}}_t)} \tag{4.42}$$

$$\propto \prod_{\substack{i=1 \\ i\neq t}}^{T} q_i(\tilde{\mathbf{s}}_i) \tag{4.43}$$

Combining the t-th factor in the likelihood function with the cavity distribution, we compute the following intermediate distribution.

$$\hat{p}_t(\tilde{S}, \mathbf{h}) = \frac{1}{Z_t} q^{\setminus t}(\tilde{S}, \mathbf{h}) p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h}) \tag{4.44}$$

where it is shown in Appendix F that the normalization constant Z_t is given by

$$Z_t = \sum_{\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_M^K} \mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t}, \Sigma_t\Big) \tag{4.45}$$

where $\Sigma_t \triangleq S_t V_h^{\setminus t} S_t^T + \frac{1}{2}\sigma_w'^2 I_{2M}$.

Since $\hat{p}_t(\tilde{S}, \mathbf{h})$ is not a member of the exponential family, it should be mapped to the

closest distribution in $\mathcal{F}$, denoted by $q^{new}(\tilde{S},\mathbf{h}) = q^{new}(\tilde{S})q^{new}(\mathbf{h})$, such that

$$q^{new}(\tilde{S},\mathbf{h}) = \underset{q(\tilde{S},\mathbf{h})\in\ \mathcal{F}}{\arg\min}\,\mathrm{KL}\Big(\hat{p}_t(\tilde{S},\mathbf{h})\|q(\tilde{S},\mathbf{h})\Big) \tag{4.46}$$

$$= \underset{q(\tilde{S},\mathbf{h})\in\ \mathcal{F}}{\arg\min}\,\mathrm{KL}\Big(\hat{p}_t(\tilde{S},\mathbf{h})\|q(\tilde{S})q(\mathbf{h})\Big). \tag{4.47}$$

It is shown in Appendix G that the above optimization problem can be divided into the following two separate optimizations:

$$q^{new}(\mathbf{h}) = \underset{q(\mathbf{h})}{\arg\min}\,\mathrm{KL}\Big(\hat{p}_t(\mathbf{h})\|q(\mathbf{h})\Big), \tag{4.48}$$

and

$$q^{new}(\tilde{S}) = \underset{q(\tilde{S})}{\arg\min}\,\mathrm{KL}\Big(\hat{p}_t(\tilde{S})\|q(\tilde{S})\Big) \tag{4.49}$$

where $\hat{p}_t(\mathbf{h})$ and $\hat{p}_t(\tilde{S})$ are the marginal distributions of $\mathbf{h}$ and $\tilde{S}$, respectively, derived from their joint distribution $\hat{p}_t(\tilde{S},\mathbf{h})$.

In the following we solve the optimizations in (4.48) and (4.49) in that order. It turns out that the solution of (4.48) is obtained from the so-called *moment matching condition* [38], and (4.49) can be solved directly.

4.4.1 Calculation of $q^{new}(\mathbf{h})$ and updating of $q_t(\mathbf{h})$

The marginal distribution of $\mathbf{h}$ is given by

$$\begin{aligned}\hat{p}_t(\mathbf{h}) &= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} \hat{p}_t(\tilde{S},\mathbf{h}) \\ &= \frac{1}{Z_t}\sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} q^{\setminus t}(\tilde{S},\mathbf{h})p(\mathbf{y}_t|\tilde{\mathbf{s}}_t,\mathbf{h}) \\ &= \frac{1}{Z_t}q^{\setminus t}(\mathbf{h})\Big(\sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} q^{\setminus t}(\tilde{S})p(\mathbf{y}_t|\tilde{\mathbf{s}}_t,\mathbf{h})\Big)\end{aligned}$$

$$= \frac{1}{Z_t} q^{\setminus t}(\mathbf{h}) \sum_{\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} p(\mathbf{y}_t | \tilde{\mathbf{s}}_t, \mathbf{h})$$

$$= \frac{1}{Z_t} q^{\setminus t}(\mathbf{h}) \psi_t(\mathbf{h}) \tag{4.50}$$

where in (4.50), $\psi_t(\mathbf{h}) \triangleq \sum_{\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} p(\mathbf{y}_t | \tilde{\mathbf{s}}_t, \mathbf{h})$, and $Z_t = \mathbb{E}_{q^{\setminus t}(\mathbf{h})}[\psi_t(\mathbf{h})]$. As expected, calculating Z_t from this gives the same result as in (4.45).

The intermediate distribution $\hat{p}_t(\mathbf{h})$ is then approximated by $q^{new}(\mathbf{h}) = \mathcal{N}(\mathbf{h} | \mathbf{m}_h^{new}, V_h^{new})$. By using the standard *assumed density filtering* (ADF) update equations [31], we compute the new moments of $q^{new}(\mathbf{h})$ as follows.

$$\mathbf{m}_h^{new} = \mathbf{m}_h^{\setminus t} + V_h^{\setminus t} \nabla_m \tag{4.51}$$

and,

$$V_h^{new} = V_h^{\setminus t} - V_h^{\setminus t} (\nabla_m \nabla_m^T - 2\nabla_v) V_h^{\setminus t} \tag{4.52}$$

where $\nabla_m \triangleq \left(\frac{\partial \log Z_t}{\partial \mathbf{m}_h^{\setminus t}} \right)^T$ and $\nabla_v \triangleq \left(\frac{\partial \log Z_t}{\partial V_h^{\setminus t}} \right)^T$ are calculated from (4.45) as follows.

$$\nabla_m = \frac{1}{Z_t} \Big(\frac{\partial \log Z_t}{\partial \mathbf{m}_h^{\setminus t}} \Big)^T$$

$$= \frac{1}{Z_t} \sum_{\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} \Big[\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t \Big) S_t^T \Sigma_t^{-1} \boldsymbol{\zeta}_t \Big] \tag{4.53}$$

and,

$$\nabla_v = \frac{1}{Z_t} \Big(\frac{\partial \log Z_t}{\partial V_h^{\setminus t}} \Big)^T$$

$$= \frac{1}{2Z_t} \sum_{\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} \Big[\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t \Big)$$

$$\times \Big(S_t^T \Sigma_t^{-1} \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t^T \Sigma_t^{-1} S_t - S_t^T \Sigma_t^{-1} S_t \Big) \Big] \tag{4.54}$$

where, $\boldsymbol{\zeta}_t \triangleq \mathbf{y}_t - S_t \mathbf{m}_h^{\setminus t}$.

After $q^{new}(\mathbf{h})$ is obtained, we can update $q_t(\mathbf{h})$ to get $q_t^{new}(\mathbf{h}) = \mathcal{N}(\mathbf{h}|\mathbf{m}_h^{new}(t), V_h^{new}(t))$ where

$$q_t^{new}(\mathbf{h}) \propto \frac{q^{new}(\mathbf{h})}{q^{\setminus t}(\mathbf{h})} = \frac{\mathcal{N}(\mathbf{h}|\mathbf{m}_h^{new}, V_h^{new})}{\mathcal{N}(\mathbf{h}|\mathbf{m}_h^{\setminus t}, V_h^{\setminus t})} \tag{4.55}$$

It is straightforward to show that,

$$V_h^{new}(t) = \Big(\big(V_h^{new}\big)^{-1} - \big(V_h^{\setminus t}\big)^{-1}\Big)^{-1} \tag{4.56}$$

and,

$$\mathbf{m}_h^{new}(t) = V_h^{new}(t)\left(\big(V_h^{new}\big)^{-1}\mathbf{m}_h^{new} - \big(V_h^{\setminus t}\big)^{-1}\mathbf{m}_h^{\setminus t}\right). \tag{4.57}$$

Note that these updates do not depend on a specific value of the transmitted vector $\tilde{\mathbf{s}}_t$, and in fact a form of averaging over all possible values of this vector appears in (4.45), (4.50), (4.53) and (4.54). □

4.4.2 Calculation of $q^{new}(\tilde{S})$ and updating of $q_t(\tilde{\mathbf{s}}_t)$

The marginal distribution of $\tilde{S}$ can be calculated as follows:

$$\begin{aligned}
\hat{p}_t(\tilde{S}) &= \int_{\mathbf{h}} \hat{p}_t(\tilde{S}, \mathbf{h})\mathrm{d}\mathbf{h} \\
&= \frac{1}{Z_t}\int_{\mathbf{h}} q^{\setminus t}(\tilde{S}, \mathbf{h}) p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h})\ d\mathbf{h} \\
&= \frac{1}{Z_t} q^{\setminus t}(\tilde{S}) \int_{\mathbf{h}} \mathcal{N}\Big(\mathbf{h}|\mathbf{m}_h^{\setminus t}, V_h^{\setminus t}\Big)\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{h}, \frac{1}{2}\sigma_w'^2 I_{2M}\Big)\mathrm{d}\mathbf{h} \\
&= \frac{1}{Z_t} q^{\setminus t}(\tilde{S})\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t}, \Sigma_t\Big) \\
&= \frac{1}{Z_t}\Big(\prod_{\substack{i=1\\ i\neq t}}^{T} q_i(\tilde{\mathbf{s}}_i)\Big)\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t}, \Sigma_t\Big)
\end{aligned} \tag{4.58}$$

Given this, the K-L divergence in (4.49) can be written as

$$\mathrm{KL}(\hat{p}_t(\tilde{S})\|q(\tilde{S})) = \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} \hat{p}_t(\tilde{S}) \log \frac{\hat{p}_t(\tilde{S})}{q(\tilde{S})}$$

$$
\begin{aligned}
&= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} \Big(\prod_{\substack{i=1 \\ i\neq t}}^{T} q_i(\tilde{\mathbf{s}}_i) \Big) \frac{\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\backslash t}, \Sigma_t\Big)}{Z_t} \\
&\qquad \times \Big[\log \frac{\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\backslash t}, \Sigma_t\Big)}{Z_t} - \log q_t(\tilde{\mathbf{s}}_t) \Big] \\
&= \sum_{\tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K} \frac{\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\backslash t}, \Sigma_t\Big)}{Z_t} \\
&\qquad \times \Big[\log \frac{\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\backslash t}, \Sigma_t\Big)}{Z_t} - \log q_t(\tilde{\mathbf{s}}_t) \Big] \qquad (4.59)
\end{aligned}
$$

The minimizer of (4.59) will be

$$
q_t^{new}(\tilde{\mathbf{s}}_t) \;= \frac{1}{Z_t} \mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\backslash t}, \Sigma_t\Big); \quad \forall \tilde{\mathbf{s}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K \qquad (4.60)
$$

Note that solving (4.49) only updates the pmf of the t-th vector, namely $\tilde{\mathbf{s}}_t$. □

Each iteration of the EP detector involves updating all the T factors of $q_t(\mathbf{h})$ using (4.56) and (4.57) and updating $q_t^{new}(s_t)$ using (4.60). After the EP algorithm converges, $\mathbf{m}_h$ is selected as the maximum likelihood (ML) estimate of the unknown channel, i.e, $\mathbf{h}' = \mathbf{m}_h$. Consequently, the channel vector in (4.15) is given by

$$
\tilde{\mathbf{h}}' = \mathbf{m}_{h,1:N_r} + j\mathbf{m}_{h,N_r+1:2N_r}. \qquad (4.61)
$$

Moreover, the following ML detection rule can be employed for the transmitted symbols

$$
\tilde{\mathbf{s}}_t' = \arg\max_{\mathbf{a} \in \mathcal{A}_{\mathcal{M}}^K} q_t(\mathbf{a}), \quad t = 1, \ldots, T. \qquad (4.62)
$$

Remark 1. *As pointed out previously, the update rules for the moments of $q_t(\boldsymbol{h})$ in* (4.56) *and* (4.57) *reveal that they are independent of the specific value of the symbol $\tilde{\mathbf{s}}_t$. In other words, refining the PDF factors for the channel is independent of refining the pmf factors for the transmitted vectors. This is to be expected since the channel vector is constant over*

the T received vectors. Consequently, it is not necessary to update the pmf of the transmitted vectors at each iteration of the EP algorithm while the moments of $q_t(\boldsymbol{h})$ are being updated using (4.56) *and* (4.57)*. In fact, calculations of* (4.60) *for all symbol vectors may be postponed to the end of the EP iterations, where a good estimate of $\boldsymbol{h}$ is at hand. Therefore, the transmitted vectors can be detected by computationally efficient linear algorithms such as zero forcing (ZF) or minimum mean-squared error (MMSE). Denoting by $\tilde{H}'$ the estimated channel matrix at the end of the EP algorithm, the MMSE estimates of the transmitted vectors in* (4.14) *are given by*

$$\tilde{\boldsymbol{s}}'_t = (\tilde{H}'^H \tilde{H}' + \sigma_w'^2 I_M / E_s)^{-1} \tilde{H}'^H \tilde{\boldsymbol{y}}_t, \quad t = 1, \ldots, T. \tag{4.63}$$

The transmitted symbols can then be detected using a demodulator which maps each component of $\tilde{\boldsymbol{s}}'_t$ to the nearest constellation point in $\tilde{\mathcal{A}}_{\mathcal{M}}$.

□

4.4.3 A Low-Complexity Approximation

A close examination of (4.45), (4.53) and (4.54) reveals that the estimation of $\mathbf{h}$ is fairly complex due to the summations involving $\mathcal{M}^K$ terms. In the following we describe a procedure which simplifies these computations. As the iterations of the EP algorithm proceed, we expect a reduction in the uncertainty regarding the channel coefficients. In other words, the mean vectors $\mathbf{m}_h(t)$ approach the actual channel vector $\mathbf{h}$, and the entries of the covariance matrices $V_h(t)$, and consequently Σ_t, become smaller. As a result, the Gaussian PDF $\mathcal{N}\left(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\right)$ will become very narrow and except for $\mathbf{y}_t$ close to its mean $S_t \mathbf{m}_h^{\setminus t}$, the function will be negligible. This implies that the values of these PDFs are negligible except for a single symbol vector. Therefore in the summations in (4.45), (4.53) and (4.54) we can ignore all the terms except for a single dominant term. To find this dominant term and the corresponding symbol vector, we assume a vectorized MIMO system similar to (4.15), in which the channel vector is given by $\mathbf{m}_h^{\setminus t}$. Now the transmission

symbol is estimated using a simple linear detector, such as ZF. Let $\tilde{\mathbf{x}}_t \in \tilde{\mathcal{A}}_{\mathcal{M}}^K$ denote the vector which maximizes $\mathcal{N}\Big(\mathbf{y}_t | S_t \mathbf{m}_h^{\setminus t}, \Sigma_t\Big)$. Then (4.53) and (4.54) can be simplified to get

$$\nabla_m \approx X_t^T \Sigma_t^{-1} \boldsymbol{\zeta}_t, \tag{4.64}$$

and

$$\nabla_v \approx X_t^T \Sigma_t^{-1} \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t^T \Sigma_t^{-1} X_t - X_t^T \Sigma_t^{-1} X_t, \tag{4.65}$$

where,

$$X_t = \begin{bmatrix} \Re(\tilde{\mathbf{x}}_t^T \otimes I_M) & -\Im(\tilde{\mathbf{x}}_t^T \otimes I_M) \\ \Im(\tilde{\mathbf{x}}_t^T \otimes I_M) & \Re(\tilde{\mathbf{x}}_t^T \otimes I_M) \end{bmatrix}. \tag{4.66}$$

Inserting the above approximations for ∇_m and ∇_v into (4.51) and (4.52), the moments of $q(\mathbf{h})$ can be updated. Furthermore, in order to improve the stability of these update rules, we employ the following smoothing mechanism [40]

$$\mathbf{m}_h^{new} = \alpha(\mathbf{m}_h^{\setminus t} + V_h^{\setminus t} \nabla_m) + (1 - \alpha)\mathbf{m}_h \tag{4.67}$$

and,

$$V_h^{new} = \alpha\big(V_h^{\setminus t} - V_h^{\setminus t}(\nabla_m \nabla_m^T - 2\nabla_v) V_h^{\setminus t}\big) + (1 - \alpha) V_h, \tag{4.68}$$

where $0 < \alpha < 1$ is a smoothing factor. After calculating $q^{new}(\mathbf{h})$, the moments of the refined version of the t-th factor are obtained as in (4.56) and (4.57).

Algorithm 3 summarizes the proposed procedure. To alleviate the ambiguity in channel estimation and to start the algorithm with a good initial value, we use the output of an EVD-based estimator to initialize the mean values of the channel coefficients, $\mathbf{m}_h(t)$.

The complexity of these operations is dominated by updating the moments of $q^{new}(\mathbf{h})$ and $q_t^{new}(\mathbf{h})$ in equations (4.56), (4.57), (4.67), and (4.68), which involve basic matrix operations such as multiplications and inversions. Accordingly, the complexity of the proposed

Data: A block of T received vectors and $p(\mathbf{y}_t|\tilde{\mathbf{s}}_t, \mathbf{h})$ distributions in (4.18)
Result: A member of exponential family as in (4.30) which is the closest PDF to (4.18), an estimation of the channel vector, and detected transmitted symbols

```
begin
    Estimate the channel with the EVD-based algorithm;
    Initialize all q_t(h) factors for t = 1, ..., T;
    Calculate q(h) using (4.33) and (4.34);
    while termination criteria has not been met do
        for t=1,...,T do
            Calculate the cavity PDF using (4.40) and (4.41);
            Estimate the maximizer vector x̃_t by m_h^{\t};
            Calculate ∇_m and ∇_v by (4.64) and (4.65), respectively;
            Find q^new(h) by (4.67) and (4.68);
            Update q_t(h) by (4.56) and (4.57);
        end
    end
    Calculate the estimated channel vector by (4.61);
    Detect the transmitted vectors using MMSE algorithm in (4.63);
end
```

Algorithm 3: Noncoherent MIMO symbol detection using expectation propagation.

algorithm can be expressed as $\mathcal{O}\big(IT(K^3M^3 + K^2M^2)\big)$, where where I denotes the number of iterations of the EP algorithm[4].

4.5 Numerical Results

In this section we investigate the channel estimation and symbol detection performances of the proposed EP algorithm, referred to as EP in the figures. We compare the result with those from EVD-based noncoherent channel estimator, which is referred to as EVD. For symbol detection we employ the MMSE algorithm for both EP and EVD receivers.

A cellular system with $L = 3$ cells, with each cell having $K = 3$ users is considered in the simulations. Without loss of generality, we consider the performance of the algorithms in the first cell. The shadowing factors for the 3 cells are chosen as $\beta_{11} = [0.98, 0.63, 0.47]$, $\beta_{12} = [0.36, 0.29, .05]$, and $\beta_{13} = [0.32, 0.14, 0.11]$.

[4]We assumed the complexity of the direct method for matrix multiplications and Gauss-Jordan elimination algorithm for matrix inversion. Therefore, the complexity of multiplying an $n \times p$ matrix by a $p \times m$ matrix is given by $\mathcal{O}(npm)$, and the complexity of inverting an $n \times n$ matrix will be $\mathcal{O}(n^3)$.

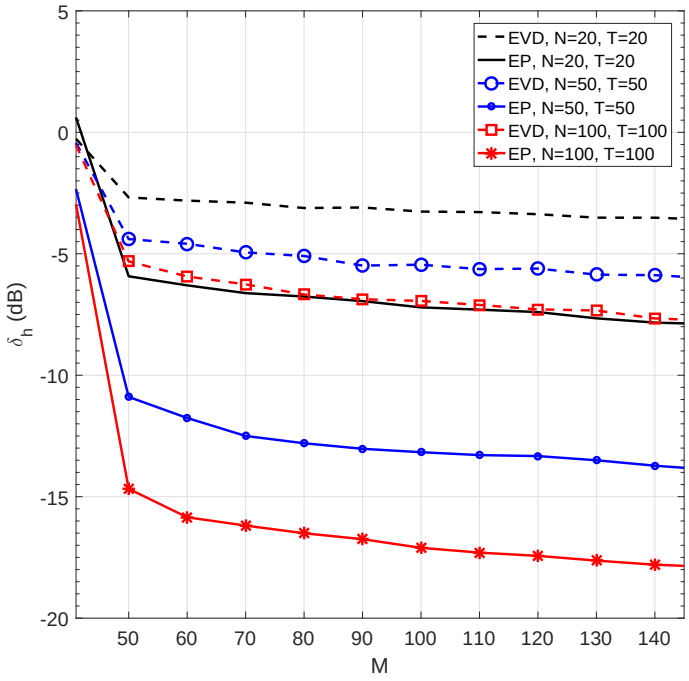


Figure 4.2: Channel estimation performance versus the receiver's antenna-array size M , for EVD and EP estimators, with symbol block sizes $N = 20, 50, 100$ and EP block size $T = N$.

We consider QPSK modulation with $E_s = 20$ dB. The EVD algorithm is assumed to work with only one pilot vector, i.e., $\tau = 1$. The EP algorithm uses the EVD's channel estimation result for initializing the mean vectors $\mathbf{m}_h(t)$. Starting with a wide search area helps in convergence of the algorithm. Therefore, we suggest initializing the covariance matrices $V_h(t)$ with fairly large values, such as $100I_{2MK}$. In our numerical results the EP algorithm was run for only 2 iterations using the smoothing factor $\alpha = 0.5$, and was applied to the entire block, i.e., $T = N$. For evaluating the channel estimation accuracy, the normalized estimation error between the channel matrix $\tilde{H}_{11}$ and its estimate $\tilde{H}'_{11}$ is considered given by $\delta_h \triangleq 10 \log_{10} \left(\|\tilde{H}_{11} - \tilde{H}'_{11}\|_F^2 / \|\tilde{H}_{11}\|_F^2 \right)$.

Remark 2. *In the multi-cell system model and considering the equivalent noise power in* (4.13)*, it is evident that interference is the dominant term in* (4.13)*. Increasing the symbol energy E_s also increases the interference and hence has a negligible effect on the signal-to-interference-plus-noise ratio (SINR). For this reason we have not shown the performance*

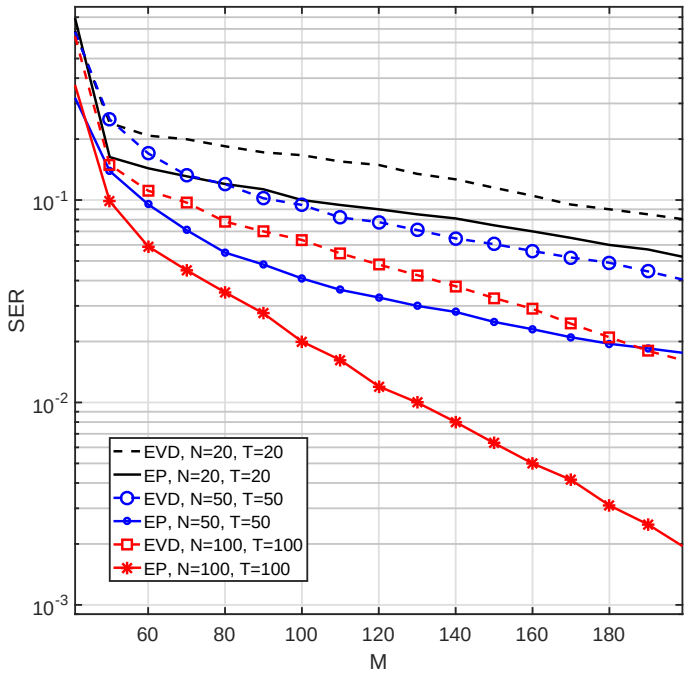


Figure 4.3: SER performance versus the receiver's antenna-array size M, for EVD and EP algorithms with symbol block sizes $N = 20, 50, 100$ and EP block size $T = N$.

of the algorithm versus SINR.

The channel estimation accuracy of the two noncohorent algorithms versus the number of receiving antennas, M, and the size of blocks, N, is depicted in Fig. 4.2. This figure shows that the performance of the two algorithms improve as M or N increases. This behavior originates from the improvements of the EVD algorithm with M and N. As discussed in Section 4.3, as M increases, the channel properties approach the favorable propagation condition, where the columns of the channel matrix become mutually orthogonal. Therefore, using larger array sizes leads to better channel estimation by EVD. On the other hand, increasing N improves the accuracy of the empirically calculated sample correlation matrix in (4.29) resulting in improved channel estimation by EVD. Since the channel estimates from EVD are used as the initial values for $\mathbf{m}_h(t)$, increases in M or N also result in improvements in the performance of the EP algorithm. Fig. 4.2 aslo shows that the EP algorithm significantly outperforms the EVD algorithm, and the improvement gain increases at higher values of M or N. For example, for $M = 100$, channel estimation

using EP improves over EVD by 7.5 dB for $N = 50$ and by 10 dB for $N = 100$.

Fig. 4.3 shows the Symbol Error Rate (SER) performance of EVD and EP algorithms versus M for three different block sizes of $N = 20, 50$ and 100 symbols. This figure also clearly shows that the proposed EP algorithm outperforms EVD. More interestingly, and similar to Fig 4.2, as the performance of EVD improves (with increases in M or N), the gain of EP over EVD also increases.

While Figs. 4.2 and 4.3 show the improvements of EP over EVD, they do not capture the entire picture. While in the case of EVD, the symbol errors are random, in the case of EP they appear in bursts. In other words, in the case of EVD, a large number of blocks will contain symbol errors whereas in the case of EP, most blocks are error free and a few blocks contain a large number of symbol error. This is demonstrated in Table 4.1 as well as in Fig. 4.4 where the percentage of symbol blocks (frames) with error is shown for both EVD and EP. According to this table, the proposed EP algorithm significantly reduces the frame error rate of the EVD algorithm. For example, for a system with $M = 150$ antennas and $N = 100$, EP reduces the percentage of erroneous frames from about 32% to 1.3%. As shown in Fig. 4.4, even for blocks as small as 10 symbols, EP reduces the frame error rate by more than 35%. As N increases, the frame error rate of the EP algorithm decreases sharply. The reason for this effect is that for the overwhelming majority of the frames where the EP algorithm converges, there are no errors in the frame. On the other hand, for some frames EP does not converge. In this case there will be a large number of

Table 4.1: The percentage of the erroneous detected frames.

	N=20		**N=50**		**N=100**	
M	**EVD**	**EP**	**EVD**	**EP**	**EVD**	**EP**
50	76.90	33.00	74.30	19.90	72.70	13.60
80	71.00	21.20	61.40	9.00	53.90	5.80
100	67.10	19.50	55.30	7.10	48.70	3.50
120	64.40	17.50	49.30	6.00	39.90	2.50
150	60.00	15.50	43.50	4.50	32.10	1.30

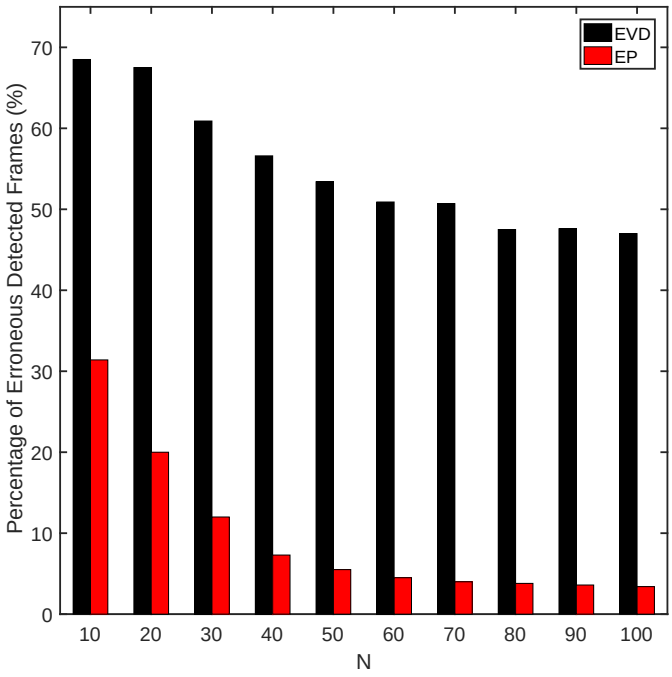


Figure 4.4: Percentage of erroneous detected frames versus symbol block size N for a MIMO system with the receiver's antenna-array size $M = 100$ and EP block size $T = N$, for EVD and EP algorithms.

symbol errors in the frame. This property of EP is very advantageous for systems using Automatic Repeat-reQuest (ARQ). The reduction in frame error rate results in significant improvements in average throughput and reduces the link delays.

The channel estimation and SER performances of the MIMO system with $M = 60$ and 100 receiving antennas versus block size N are depicted in Figs. 4.5 and 4.6, respectively. Again, these figures show the improvements of EP over EVD. Moreover, as the performance of EVD improves, the gain of EP over EVD also increases.

Fig. 4.7 shows the effectiveness of the EP algorithm in combating inter-cell interference. The SER performances of the two algorithms versus M are shown for a multi-cell and a single-cell system. Comparing the performance of EVD for single-cell and multi-cell systems clearly shows the adverse effect of intra-cell interference. In contrast, the EP algorithm shows only a minor degradation in performance due to the additional intra-cell interference.

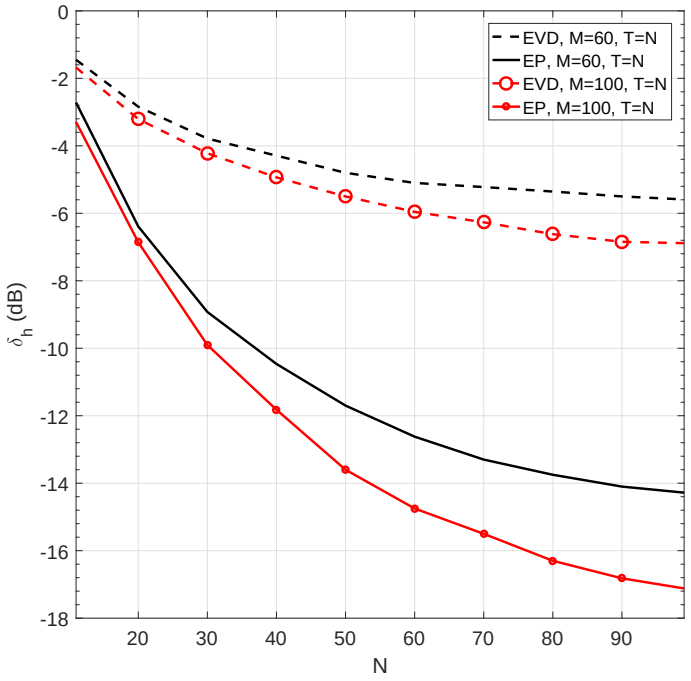


Figure 4.5: Channel estimation performance of EVD and EP algorithms versus symbol block size N for a MIMO system with the receiver's antenna-array sizes M =60,100, and EP block size $T = N$.

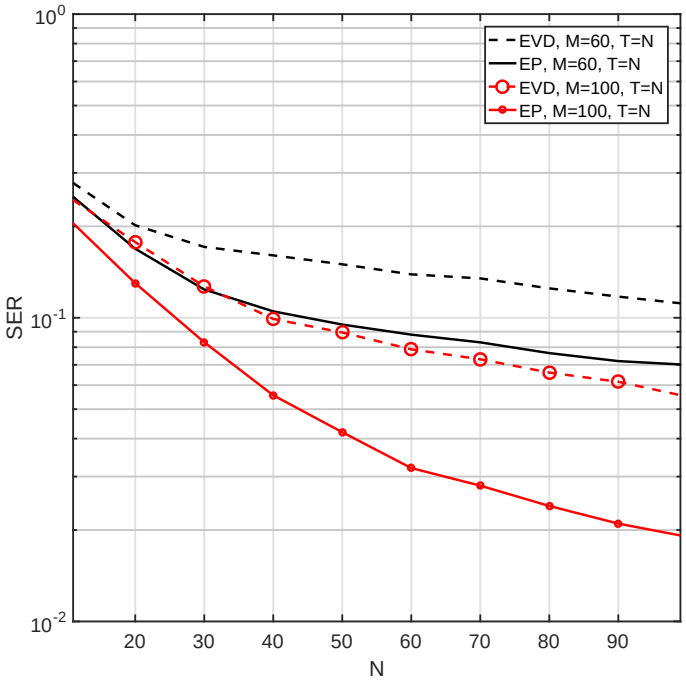


Figure 4.6: SER performance of EVD and EP algorithms versus symbol block size N for a MIMO system with the receiver's antenna-array sizes M =60,100, and EP block size $T = N$

.

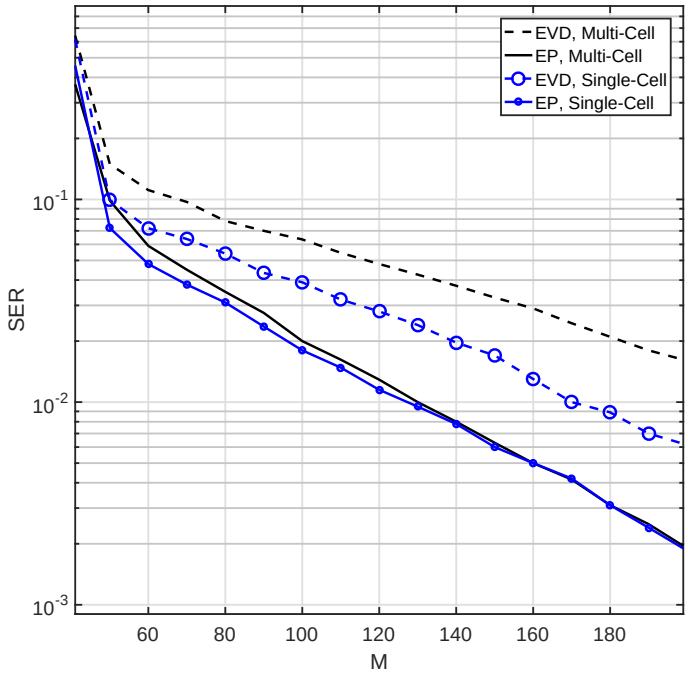


Figure 4.7: Multi-cell versus single-cell performances for EVD and EP detectors.

Chapter 5
Conclusion

In Chapter 2 we investigate the performance of the expectation propagation (EP) detector in practical situations in which perfect channel state information (CSI) is not available at the receiver. As expected and verified by the simulation results, lack of perfect CSI results in significant performance loss for the EP detector. Moreover, the EP detector shows a higher sensitivity to the channel estimation error at high signal-to-noise ratios (SNR) and the rate of its performance improvement decreases at higher SNRs. To rectify this problem we propose a Modified EP algorithm for correlated noise which utilizes the correlation matrix of the channel estimation error. Simulation results show that the modified algorithm is robust against imperfect CSI and its performance is significantly improved over the EP algorithm.

A noncoherent detector for large-scale SIMO systems using the Expectation Propagation algorithm is proposed in Chapter 3. We show through simulation that with only a single iteration and for block sizes as small as two symbols, the proposed algorithm outperforms the pilot-based MMSE detector in term of symbol error rate (SER). This property makes the proposed algorithm suitable for channels with coherence times as short as two symbol durations. Moreover, the simulation results verify that as the symbol block size increases, the SER performance of the algorithm converges to that of the optimal ML receiver which has perfect knowledge of channel state information. The proposed detector does not rely on prior knowledge of channel statistics and it is shown that for a Rician fading channel, it can outperform the coherent MMSE detector without using Rician K-factor. Finally, the proposed method does not rely on specific signal constellations and can be used for any differential modulation scheme.

In Chapter 4 we propose a noncoherent channel estimation and symbol detection algorithm for multi-cell multi-user massive MIMO systems based on the expectation propa-

gation algorithm. The proposed algorithm is initialized with the channel estimation result from the EVD-based method. Simulation results show that after a few iterations, the EP-based algorithm significantly outperforms the EVD-based method in both channel estimation and symbol error rate. Moreover, the EP-based algorithm is not sensitive to antenna array size or the inaccuracies of sample correlation matrix.

References

[1] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. wall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 287–293.

[2] the network, cisco's technology news site. [Online]. Available: https://newsroom.cisco.com/press-release-content?articleId=1741352#_ftn1.

[3] cisco. [Online]. Available: http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html.

[4] 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, Autumn 1996.

[5] E. Telatar, "Capacity of multi-antenna gaussian channels." *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999. [Online]. Available: http://dblp.uni-trier.de/db/journals/ett/ett10.html#Telatar99

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas." *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998. [Online]. Available: http://dblp.uni-trier.de/db/journals/wpc/wpc6.html#FoschiniG98

[7] J. Hampton, *Introduction to MIMO communications*. Cambridge: Cambridge University Press, 2014.

[8] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[10] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 742–758, Oct 2014.

[11] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[12] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *Information Theory, IEEE Transactions on*, vol. 50, no. 9, pp. 1893–1909, Sept 2004.

[13] C. Zhang and R. Qiu, "Massive MIMO as a big data system: Random matrix models and testbed," *Access, IEEE*, vol. 3, pp. 837–851, 2015.

[14] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *Globecom Workshops (GC Wkshps), 2014*, Dec 2014, pp. 287–293.

[15] P. Harris, S. Zang, A. Nix, M. Beach, S. Armour, and A. Doufexi, "A distributed massive MIMO testbed to assess real-world performance and feasibility," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, May 2015, pp. 1–2.

[16] A. Tulino and S. Verdu, *Random matrix theory and wireless communications*. Delft, The Netherlands: Now Publishers, Inc., 2004.

[17] A. Chockalingam and B. Rajan, *Large MIMO Systems*, ser. Large MIMO Systems. Cambridge University Press, 2014.

[18] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct 2003.

[19] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[20] C. Kong, C. Zhong, A. K. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Effect of channel aging on the sum rate of uplink massive MIMO systems," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1222–1226.

[21] G. Amarasuriya and H. Poor, "Impact of channel aging in multi-way relay networks with massive MIMO," in *Communications (ICC), 2015 IEEE International Conference on*, June 2015, pp. 1951–1957.

[22] K. Truong and R. Heath, "Effects of channel aging in massive MIMO systems," *Communications and Networks, Journal of*, vol. 15, no. 4, pp. 338–351, Aug 2013.

[23] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *Communications, IEEE Transactions on*, vol. 54, no. 3, pp. 543–552, March 2006.

[24] J. Proakis and M. Salehi, *Digital Communications*. New York: McGraw-Hill, 2008.

[25] 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.

[26] G. Ginis and J. Cioffi, "On the relation between v-blast and the gdfe," *Communications Letters, IEEE*, vol. 5, no. 9, pp. 364–366, Sept 2001.

[27] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the ul/dl of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, February 2013.

[28] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[29] J. Goldberger and A. Leshem, "MIMO detection for high-order qam based on a gaussian tree approximation," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4973–4982, Aug 2011.

[30] J. Goldberger, "Improved MIMO detection based on successive tree approximations," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 2004–2008.

[31] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[32] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2840–2849, Aug 2014.

[33] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[34] T. Bogale and L. B. Le, "Pilot optimization and channel estimation for multiuser massive MIMO systems," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, March 2014, pp. 1–6.

[35] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive MIMO using gaussian-mixture bayesian learning," *Wireless Communications, IEEE Transactions on*, vol. 14, no. 3, pp. 1356–1368, March 2015.

[36] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5795–5809, Oct 2015.

[37] B. S. Thian and A. Goldsmith, "Decoding for MIMO systems with imperfect channel state information," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, Dec 2010, pp. 1–6.

[38] T. Minka, "A family of algorithms for approximate Bayesian inference," PhD thesis, MIT, 2001. [Online]. Available: http://research.microsoft.com/ minka/papers/ep/minka-thesis.pdf

[39] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.

[40] M. Seeger, "Expectation propagation for exponential families," University of California, Berkeley, CA, USA, Tech. Rep., 2005.

[41] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, Aug 2002.

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

[43] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (corresp.)," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 619–621, October 1967.

[44] T. L. Marzetta, "How much training is required for multiuser mimo?" in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 359–363.

[45] M. Stojnic and B. Hassibi, "Out-sphere decoder for non-coherent ml SIMO detection and its expected complexity," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, Nov 2007, pp. 1568–1572.

[46] H. A. J. Alshamary, T. Al-Naffouri, A. Zaib, and W. Xu, "Optimal non-coherent data detection for massive SIMO wireless systems: A polynomial complexity solution," in *Signal Processing and Signal Processing Education Workshop (SP/SPE), 2015 IEEE*, Aug 2015, pp. 172–177.

[47] M. Chowdhury, A. Manolakos, and A. J. Goldsmith, "Design and performance of noncoherent massive simo systems," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, March 2014, pp. 1–6.

[48] B. Knott, M. Chowdhury, A. Manolakos, and A. J. Goldsmith, "Benefits of coding in a noncoherent massive SIMO system," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 2350–2355.

[49] A. Manolakos, M. Chowdhury, and A. J. Goldsmith, "Constellation design in noncoherent massive SIMO systems," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 3690–3695.

[50] A. G. Armada and L. Hanzo, "A non-coherent multi-user large scale SIMO system relaying on m-ary dpsk," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 2517–2522.

[51] Y. Qi and T. P. Minka, "Window-based expectation propagation for adaptive signal detection in flat-fading channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 348–355, Jan 2007.

[52] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 905–923, Secondquarter 2016.

[53] K. Ghavami and M. Naraghi-Pour, "Noncoherent SIMO detection by expectation propagation," in *2017 IEEE International Conference on Communications (ICC)*, 2017.

[54] H. Q. Ngo and E. G. Larsson, "Evd-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3249–3252.

[55] L. Cottatellucci, R. R. Müller, and M. Vehkaperä, "Analysis of pilot decontamination based on power control," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[56] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, Oct 2014.

[57] R. R. Müller, M. Vehkaperä, and L. Cottatellucci, "Analysis of blind pilot decontamination," in *2013 Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 1016–1020.

[58] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3421–3425.

[59] C. K. Wen, S. Jin, K. K. Wong, J. C. Chen, and P. Ting, "Channel estimation for massive MIMO using gaussian-mixture bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1356–1368, March 2015.

[60] K. Ghavami and M. Naraghi-Pour, "MIMO detection with imperfect channel state information using expectation propagation," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[61] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Exploring Artificial Intelligence in the New Millennium*, G. Lakemeyer and B. Nebel, Eds. Morgan Kaufmann Publishers, Jan. 2003, ch. 8, pp. 239–236. [Online]. Available: http://www.merl.com/publications/TR2001-22

[62] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[63] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[64] J. Hu and T. M. Duman, "Graph-based detector for blast architecture," in *2007 IEEE International Conference on Communications*, June 2007, pp. 1018–1023.

Appendix A
Statistical Inference by Message Passing

In this appendix we briefly introduce some useful graphs and the sum-product algorithm over them. To show the applicability of the algorithms in digital communications, two simple examples in MIMO detection problem are presented.

A.1 Graphical Structures

In general, graph-based statistical inference algorithms start with modeling the joint probability distributions with a graphical model. Factor graphs and Markov random fields, as two common and useful graphs for statistical inference problems, will be introduced in following sections.

Factor Graphs: Factor graphs[1] consist of two types of nodes (or vertices), namely, variable nodes and factor nodes. What makes these graphs distinguishable is that their graph edges are only connected between a variable node and a factor node. Any factorisable joint PDF can be modeled with a factor graph. For example, suppose the joint PDF of four random variables can be factorized as $p(x_1, x_2, x_3, x_4) = f_1(x_1)f_2(x_1, x_4)f_3(x_2, x_3, x_4)$. Then, as depicted in Fig. A.1, this PDF can be modeled with a factor graph consists of four variable nodes and three factor nodes

Accordingly, the general form of a factorisable joint distribution of n random variables, such as $x_1, \ldots, x_n$, that can be modeled with a factor graph is as follows

$$p(x_1, \ldots, x_n) = \prod_j f_j(X_j) \tag{A.1}$$

in which, X_j is a subset of the set of random variables.

Markov Random Fields: A Markov random field (MRF) is a graph structure in

[1] Also known as bipartite graphs.

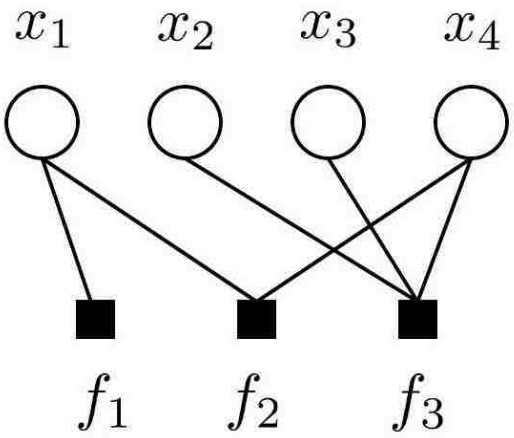


Figure A.1: A sample factor graph representing $f_1(x_1)f_2(x_1, x_4)f_3(x_2, x_3, x_4)$.

which the conditional distribution of each node, given its neighbors, is independent of the other nodes in the graph. In other words,

$$p(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = p(x_i|\mathcal{N}(x_i)) \tag{A.2}$$

in which, the neighbouring nodes of x_i are denoted by $\mathcal{N}(x_i)$.

A fully-connected subgraph in a MRF is known as a *clique*. A MRF with the maximum clique size of two is called *pairwise MRF*. One useful pairwise MRF for the applications of statistical inference is shown in Fig. A.2. As can be seen, the nodes in this type of graph can be categorized in the observed variables (shown by filled circles) and hidden variables. The conditional distribution of the observed variables, given it adjacent hidden variable, is independent of all other hidden nodes. Also, each hidden node follows the Markovian property given in (A.2). Suppose the hidden nodes and observed nodes are denoted by x_is and y_is, respectively. It is common to model the dependency between an observed node like and its corresponding hidden node with a function like $\phi_i(x_i, y_i)$. Also, the dependency between each hidden node and its neighbors are modeled with compatibility functions as $\psi_{ij}(x_i, x_j)$. Consequently, the joint PDF of hidden and observed variables in a pairwise

MRF can be represented as [61]

$$p(x_1, \ldots, x_n, y_1, \ldots, y_n) = \frac{1}{Z} \prod_{i,j} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i, y_i), \tag{A.3}$$

in which, Z is a normalization constant.

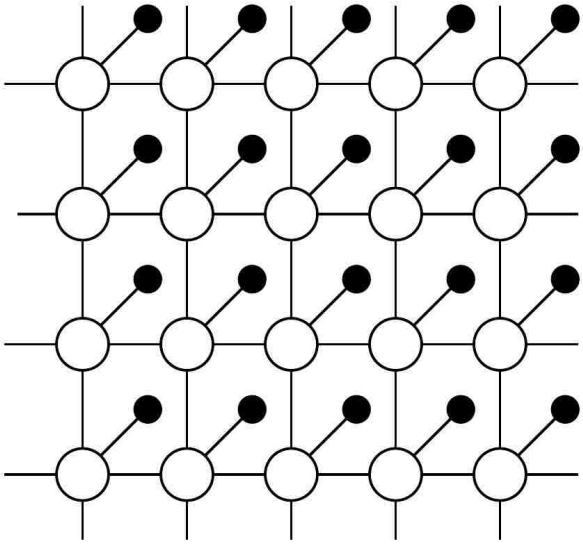


Figure A.2: A sample pairwise MRF.

Since the observations are fixed during the inference process, they can be omitted in the joint PDF. Therefore a pairwise MRF can also be modeled as [61]

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{i,j} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i) \tag{A.4}$$

A.2 Sum-Product Algorithm

The sum-product or belief propagation (BP) is an algorithm for calculating the marginal PDF of variables by iteratively exchanging messages or beliefs between the nodes in a graph [62, 61, 63]. This algorithm can be applied to both graphical structures we introduced earlier.

In a factor graph, the sum-product algorithm needs two types of messages, one from a variable node to a factor node and the other type of message for the reverse direction.

Following the well-known notation in [62], the sum-product messages from a variable node x to a factor node f, and vice versa, in a factor graph are given as follows

$$\mu_{x\to f}(x) = \prod_{h\in\mathcal{N}(x)\setminus\{f\}} \mu_{h\to x}(x) \tag{A.5}$$

and

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} f(X) \prod_{y\in\mathcal{N}(f)\setminus\{x\}} \mu_{y\to f}(y) \tag{A.6}$$

where, $\mathcal{N}(x)$ shows the neighbours of x, $\sim\{x\}$ means all variables but x, and the backslash signes in $\setminus\{f\}$ and $\setminus\{x\}$ simply mean excluding the factor node f or variable node x, respectively. After algorithm's convergence, the marginal distribution of each variable node can be calculated as

$$f(x) = \prod_{h\in\mathcal{N}(x)} \mu_{h\to x}(x) \tag{A.7}$$

The sum-product algorith can be implemented in pairwise MRF by exchanging the following message between the hidden nodes[61]:

$$m_{ij}(x_j) = \sum_{x_i} \phi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{k\in\mathcal{N}(i)\setminus j} m_{ki}(x_i). \tag{A.8}$$

which can also be considered as the believe of the i-th hidden node in the value of the j-th hidden node. Finally, the belief of the i-th hidden node is given as

$$b_i(x_i) = k\phi_i(x_i) \prod_{j\in\mathcal{N}(i)} m_{ji}(x_i). \tag{A.9}$$

A.3 Applications of the BP Algorithm in MIMO Detection

To show the applicability of the BP algorithm to digital communications, in this section we present two examples of MIMO detection by this algorithm.

Assume a MIMO system with M receiving and K transmitting antennas. The received vector for this model is given as

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}, \tag{A.10}$$

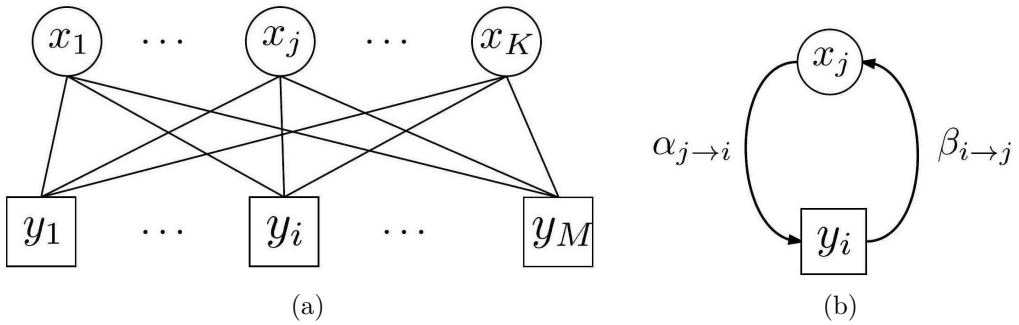
in which, $\mathbf{y} \in \mathbb{C}^{M\times 1}$ is the received vector, $\mathbf{x} \in \mathbb{C}^{K\times 1}$ is the vector of K transmitted symbols as $\mathbf{x} = [x_1, \ldots, x_K]^T$, $n \in \mathbb{C}^{M\times 1}$ is the AWGN noise vector with distribution $\mathcal{CN}(0, \sigma^2 I_M)$, and $H \in \mathbb{C}^{M\times K}$ is the known channel matrix.

MIMO detection can be considered as a statistical inference problem. In fact, the symbol detection is equivalent to find the marginal PDF of the transmitted vectors. Therefore, as we will show in following parts, by assigning the transmitted symbols to the variable nodes of a factor graph or the hidden nodes of a pairwise MRF, it is possible to iteratively find their marginals by BP algorithm.

MIMO Detection by Factor Graph Model: As is depicted in Fig.A.3, the MIMO detection problem can be modeled with a factor graph where its variable nodes are the transmitted symbols and its factor nodes are the received signals [64]. This leads to a densely connected graph with lots of cycles which affect the detection performance and accuracy of the detector.

By assuming BPSK symbols, the log likelihood ratio of the j-th transmitted symbol given the i-th received signal is given as

$$\log \frac{p(x_j = 0|y_i)}{p(x_j = 1|y_i)} = \log \frac{\sum_{\mathbf{x}, x_j=0} p(\mathbf{x}|y_i)}{\sum_{\mathbf{x}, x_j=1} p(\mathbf{x}|y_i)}$$



(a) (b)

Figure A.3: Graphical model for MIMO detection. (a) Factor Graph, (b) Messages.

$$\begin{aligned} &= \log \frac{\sum_{\mathbf{x},x_j=0} p(\mathbf{x})p(y_i|\mathbf{x})}{\sum_{\mathbf{x},x_j=1} p(\mathbf{x})p(y_i|\mathbf{x})} \\ &= \log \frac{\sum_{\mathbf{x},x_j=0} p(x_1,\ldots,x_K)p(y_i|\mathbf{x})}{\sum_{\mathbf{x},x_j=1} p(x_1,\ldots,x_K)p(y_i|\mathbf{x})} \\ &= \log \frac{p(x_j=0)}{p(x_j=1)} + \log \frac{\sum_{\mathbf{x},x_j=0} p(y_i|\mathbf{x}) \prod_{k\neq j} p(x_k)}{\sum_{\mathbf{x},x_j=1} p(y_i|\mathbf{x}) \prod_{k\neq j} p(x_k)}. \end{aligned} \tag{A.11}$$

By defining $\alpha_j \triangleq \log \frac{p(x_j=0)}{p(x_j=1)}$, we can write

$$p(x_j) = \frac{e^{-\alpha_j}}{1+e^{-\alpha_j}} e^{\alpha_j x_j} \tag{A.12}$$

By denoting the i-th row of the channel matrix by $\mathbf{h}_i$, the i-th received signal can be written as $y_i = \mathbf{h}_i\mathbf{x} + n_i$. Therefore,

$$p(y_i|\mathbf{x}) = \frac{1}{\pi\sigma^2} \exp\big(-\frac{1}{\sigma^2}|y_i - \mathbf{h}_i\mathbf{x}|^2\big). \tag{A.13}$$

Therefore, the log likelihood ration in (A.11) can be written as

$$\log \frac{p(x_j=0|y_i)}{p(x_j=1|y_i)} = \alpha_j + \beta_{i\to j} \tag{A.14}$$

in which,

$$\beta_{i\to j} = \log \frac{\sum_{\mathbf{x},x_j=0} \exp\big(-\frac{1}{\sigma^2}|y_i - \mathbf{h}_i\mathbf{x}|^2 + \sum_{k\neq j}\alpha_k x_k\big)}{\sum_{\mathbf{x},x_j=1} \exp\big(-\frac{1}{\sigma^2}|y_i - \mathbf{h}_i\mathbf{x}|^2 + \sum_{k\neq j}\alpha_k x_k\big)}. \tag{A.15}$$

According to the sum-product algorithm, the message from j-th variable node to the

i-th factor node is given as

$$\alpha_{j\to i} = \sum_{m=1, m\neq i}^{M} \beta_{m\to j} \tag{A.16}$$

Finally, the belief of the j-th transmitted symbol can be calculated as

$$L(x_j) = \sum_{m=1}^{M} \beta_{m\to j} \tag{A.17}$$

The detector can use the $L(.)$ values to decide about the transmitted symbols.

MIMO Detection by Markov Random Field Model: To fit the MIMO detection problem to a pairwise MRF, first we have to transform the joint distribution of hidden nodes into the structure given in (A.4). The joint conditional PDF of the trasnmited symbols is given as follows [29]

$$\begin{aligned} P(\mathbf{x}|\mathbf{y}) &= P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\ &= \frac{1}{(\pi\sigma^2)^M} e^{-\frac{1}{\sigma^2}(\mathbf{y}-H\mathbf{x})^H(\mathbf{y}-H\mathbf{x})} \prod_{i=1}^{K} P(x_i) \\ &\propto \exp\left(-\frac{1}{\sigma^2}(\mathbf{x}^H H^H H\mathbf{x} - \mathbf{y}^H H\mathbf{x} - \mathbf{x}^H H^H \mathbf{y})\right) \prod_{i=1}^{K} \exp(\ln P(x_i)) \\ &\propto \exp\left(-\frac{1}{\sigma^2}[\mathbf{x}^H H^H H\mathbf{x} - 2\Re(\mathbf{x}^H H^H \mathbf{y})]\right) \prod_{i=1}^{K} \exp(\ln P(x_i)). \end{aligned} \tag{A.18}$$

By assuming $R = H^H H$ and $\mathbf{z} = H^H \mathbf{y}$, the above expression can be further simplified as

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{\sigma^2}[\mathbf{x}^H R\mathbf{x} - 2\Re(x^H \mathbf{z})]\right) \prod_{i=1}^{K} \exp(\ln P(x_i)) \tag{A.19}$$

Also,

$$\begin{aligned}
\mathbf{x}^H R\mathbf{x} &= \begin{bmatrix} x_1^* \dots x_K^* \end{bmatrix} \begin{bmatrix} r_{11} & \dots & r_{1K} \\ \vdots & \ddots & \vdots \\ r_{K1} & \dots & r_{KK} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix} \\
&= \sum_{i=1}^{K} \sum_{j=1}^{K} x_i^* r_{ij} x_j \\
&= \sum_{i=1}^{K} \sum_{j=1, j\neq i}^{N_T} x_i^* r_{ij} x_j + \sum_{i=1}^{K} |x_i|^2 r_{ii} \\
&= \sum_{i=1}^{K} \sum_{j<i} 2\Re(x_i^* r_{ij} x_j) + \sum_{i=1}^{K} |x_i|^2 r_{ii} \qquad \text{(A.20)}
\end{aligned}$$

and,

$$\begin{aligned}
\mathbf{x}^H \mathbf{z} &= \begin{bmatrix} x_1^* \dots x_K^* \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_K \end{bmatrix} \\
&= \sum_{i=1}^{K} x_i^* z_i \qquad \text{(A.21)}
\end{aligned}$$

Finally,

$$\begin{aligned}
P(\mathbf{x}|\mathbf{y}) &\propto \exp\left(-\frac{1}{\sigma^2} [\sum_{i=1}^{K} \sum_{j<i} 2\Re(x_i^* r_{ij} x_j) + \sum_{i=1}^{K} |x_i|^2 r_{ii} - 2\sum_{i=1}^{K} \Re(x_i^* z_i)] \right) \\
&\quad \times \prod_{i=1}^{K} \exp(\ln P(x_i)) \\
&\propto \exp\left(-\frac{2}{\sigma^2} \sum_{i=1}^{K} \sum_{j<i} \Re(x_i^* r_{ij} x_j) \right) \exp\left(-\frac{1}{\sigma^2} [\sum_{i=1}^{K} |x_i|^2 r_{ii} - 2\sum_{i=1}^{K} \Re(x_i^* z_i)] \right) \\
&\quad \times \prod_{i=1}^{K} \exp(\ln P(x_i)) \\
&\propto \prod_{i=1}^{K} \prod_{j<i} \exp\left(-\frac{2}{\sigma^2} \Re(x_i^* r_{ij} x_j) \right)
\end{aligned}$$

$$\times \prod_{i=1}^{K} \exp\left(-\frac{1}{\sigma^2}[|x_i|^2 r_{ii} - 2\Re(x_i^* z_i)] + \ln P(x_i)\right) \tag{A.22}$$

Comparing (A.22) with (A.4) reveals that

$$\psi_{ij}(x_i, x_j) = \exp\left(-\frac{2}{\sigma^2}\Re(x_i^* r_{ij} x_j)\right) \tag{A.23}$$

and,

$$\phi_i(x_i) = \exp\left(-\frac{1}{\sigma^2}[|x_i|^2 r_{ii} - 2\Re(x_i^* z_i)] + \ln P(x_i)\right) \tag{A.24}$$

By these functions, the receiver can calculate the messages given in (A.8). After enough number of iterations, the receiver can finally calculate the belief of each transmitted symbol by (A.9).

Appendix B
Properties of Gaussin Random Vectors

The following two properties are widely used in finding the marginal and conditional distributions of Gaussian random vectors.

A.1

Suppose $\mathbf{x} \sim \mathcal{N}(\mathbf{x}|\mathbf{m}_x, V_x)$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{n}|\mathbf{0}, V_n)$ are two independent Gaussian random vectors. Therefore, the joint PDF of $\mathbf{x}$ and $\mathbf{y} = H\mathbf{x} + \mathbf{n}$ is given as

$$\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\middle| \begin{pmatrix}\mathbf{m}_x\\ H\mathbf{m}_x\end{pmatrix}, \begin{pmatrix} V_x & V_xH^T \\ HV_x & HV_xH^T + V_n\end{pmatrix}\right) \tag{B.1}$$

A.2

Assume the joint PDF of two random vectors $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ as follows

$$\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\middle| \begin{pmatrix}\mathbf{m}_x\\ \mathbf{m}_y\end{pmatrix}, \begin{pmatrix} V_x & V_{xy} \\ V_{xy}^T & V_y\end{pmatrix}\right) \tag{B.2}$$

Then the marginal and conditional distributions of the two variables are given as

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}|\mathbf{m}_x, V_x) \tag{B.3}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{y}|\mathbf{m}_y, V_y) \tag{B.4}$$

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}\Big(\mathbf{x}|\mathbf{m}_x + V_{xy}V_y^{-1}(\mathbf{y} - \mathbf{m}_y), V_x - V_{xy}V_y^{-1}V_{xy}^T\Big) \tag{B.5}$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}\Big(\mathbf{y}|\mathbf{m}_y + V_{xy}^TV_x^{-1}(\mathbf{x} - \mathbf{m}_x), V_y - V_{xy}^TV_x^{-1}V_{xy}\Big) \tag{B.6}$$

Appendix C
Assumed Density Filtering

Assume x as the latent variable and y_t as the measured sample at time index t. Therefore the measurements in a duration of T time indexes can be represented by the vector

$$\mathbf{y}_{1:T} = (y_1, \ldots, y_t, y_{t+1}, \ldots, y_T). \tag{C.1}$$

We also consider the conditionally independent measurements for which the current measurement y_t given the latent variable is independent of all previous $t-1$ measurements, i.e. $p(y_t|x, \mathbf{y}_{1:t-1}) = p(y_t|x)$.

Generally, the two probability density functions (PDF) $p(x|\mathbf{y}_{1:T})$ and $p(\mathbf{y}_{1:T})$ are important for estimation of the latent variable and system modeling, respectively. The aim in the online Bayesian inference methodology is to calculate these distributions iteratively after each measurement. Therefore, instead of calculating $p(x|\mathbf{y}_{1:T})$ after collecting all T measurements, the online algorithm calculates $p(x|\mathbf{y}_{1:t})$ by receiving t-th measurement y_t, as follows

$$\begin{aligned} p(x|\mathbf{y}_{1:t}) &= \frac{p(x, \mathbf{y}_{1:t})}{P(\mathbf{y}_{1:t})} \\ &= \frac{p(x, \mathbf{y}_{1:t-1}, y_t)}{p(\mathbf{y}_{1:t-1}, y_t)} \\ &= \frac{p(x, \mathbf{y}_{1:t-1})p(y_t|x, \mathbf{y}_{1:t-1})}{p(\mathbf{y}_{1:t-1})p(y_t|\mathbf{y}_{1:t-1})} \\ &= \frac{p(x|\mathbf{y}_{1:t-1})p(y_t|x)}{p(y_t|\mathbf{y}_{1:t-1})} \end{aligned} \tag{C.2}$$

Therefore,

$$p(x|\mathbf{y}_{1:t}) = \frac{1}{Z_t} p(y_t|x) p(x|\mathbf{y}_{1:t-1}) \tag{C.3}$$

in which, the constant Z_t is defined as $Z_t \triangleq p(y_t|\mathbf{y}_{1:t-1})$. Equation (C.3) is used in the

update step of Bayesian inference algorithms, like Kalman filters. Considering the definition of Z_t, $p(\mathbf{y}_{1:T})$ can be calculated as follows

$$\begin{aligned} p(\mathbf{y}_{1:T}) &= p(y_1, \ldots, y_T) && \text{(C.4)} \\ &= p(y_1)p(y_2|y_1)\ldots p(y_T|y_1, \ldots, y_{T-1}) && \text{(C.5)} \\ &= Z_1 Z_2 \ldots Z_T && \text{(C.6)} \end{aligned}$$

One big problem in (C.3) is its dependency to all $t-1$ previous measurements. As t grows up, the complexity of calculating and handling $p(x|\mathbf{y}_{1:t-1})$ increases. One subtle way to solve this problem is using an approximation of $p(x|\mathbf{y}_{1:t-1})$ which is more tractable. In assumed density filtering (ADF) algorithm, $p(x|\mathbf{y}_{1:t-1})$ is approximated by a member of the exponential family of distributions, denoted by $\mathcal{F}$, as $q^{\backslash t}(x)$. Consequently, (C.3) can be written as the following intermediate distribution

$$\hat{p}(x|\mathbf{y}_{1:t}) = \frac{1}{Z_t} p(y_t|x) q^{\backslash t}(x) \tag{C.7}$$

in which, $Z_t \triangleq \mathbb{E}_{q^{\backslash t}}[p(y_t|x)]$. Since $\hat{p}(x|\mathbf{y}_{1:t})$ generally is not a member of the exponential family, it should be approximated by a PDF like $q^{\backslash t+1}(x)$ from this family such that satisfies the following optimization problem

$$q^{\backslash t+1}(x) = \underset{q(x)\in\mathcal{F}}{\arg\min}\, \mathbb{KL}\Big(\hat{p}(x|\mathbf{y}_{1:t})||q(x)\Big) \tag{C.8}$$

In the other word, the approximating PDF has the minimum Kullback-Leibler divergence to the exact distribution. We will see that the solution of this optimization problem can be easily found by the so called *moment matching* property.

Appendix D
Proof of the Moment Matching Property

Any member of the family of exponential distribution, like $q(x)$, can be represented as

$$q(x) = h(x)g(\boldsymbol{\eta}) \exp\left(\boldsymbol{\eta}^T \mathbf{u}(x)\right) \tag{D.1}$$

in which $\mathbf{u}(x)$ is the vector of sufficient statistics and $\boldsymbol{\eta}$ is the vector of natural parameters. $g^{-1}(\boldsymbol{\eta})$ is the partition function and can be calculated as

$$g(\boldsymbol{\eta}) = \frac{1}{\int_x h(x) \exp\left(\boldsymbol{\eta}^T \mathbf{u}(x)\right) \mathrm{d}x}. \tag{D.2}$$

And its logarithm is given as

$$\log g(\boldsymbol{\eta}) = -\log \int_x h(x) \exp\left(\boldsymbol{\eta}^T \mathbf{u}(x)\right) \mathrm{d}x. \tag{D.3}$$

By calculating the gradient of (D.3) with respect to the vector of natural parameters, we can write

$$\begin{aligned}
\nabla_{\boldsymbol{\eta}} \log g(\boldsymbol{\eta}) &= -\frac{1}{\int_x h(x) \exp\left(\boldsymbol{\eta}^T \mathbf{u}(x)\right) \mathrm{d}x} \nabla_{\boldsymbol{\eta}} \int_x h(x) \exp\left(\boldsymbol{\eta}^T \mathbf{u}(x)\right) \mathrm{d}x \\
&= -g(\boldsymbol{\eta}) \int_x h(x) \exp\left(\boldsymbol{\eta}^T \mathbf{u}(x)\right) \mathbf{u}(x) \mathrm{d}x \\
&= -\int_x q(x) \mathbf{u}(x)\big) \mathbf{u}(x) \mathrm{d}x \\
&= -\mathbb{E}_q[\mathbf{u}(x)].
\end{aligned} \tag{D.4}$$

Suppose we are interested in the solution of the following optimization problem:

$$q^*(x) = \arg\min_{q(x)\in\mathcal{F}} \mathbb{KL}\Big(p(x)||q(x)\Big) \tag{D.5}$$

where $p(x)$ is an arbitrary PDF not necessarily in $\mathcal{F}$. According to the definition of KL divergence, the expression under optimization can be expanded as follows[1]

$$\begin{aligned}
\mathbb{KL}\Big(p(x)||q(x)\Big) &= \int_x p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \\
&= \int_x p(x)\log p(x)\mathrm{d}x - \int_x p(x)\log\Big(h(x)g(\boldsymbol{\eta})\exp(\boldsymbol{\eta}^T\mathbf{u}(x))\Big)\mathrm{d}x \\
&= -\int_x p(x)\log g(\boldsymbol{\eta})\mathrm{d}x - \int_x p(x)\boldsymbol{\eta}^T\mathbf{u}(x)\mathrm{d}x + \mathrm{C} \\
&= -\log g(\boldsymbol{\eta}) - \boldsymbol{\eta}^T\mathbb{E}_p[\mathbf{u}(x)] + \mathrm{C}
\end{aligned} \tag{D.6}$$

in which, C is a constant value with respect to $\boldsymbol{\eta}$. By setting the gradient of this statement with respect to $\boldsymbol{\eta}$ to zero,

$$\mathbb{E}_p[\mathbf{u}(x)] = -\frac{1}{g(\boldsymbol{\eta})}\nabla_{\boldsymbol{\eta}}g(\boldsymbol{\eta}) = -\nabla_{\boldsymbol{\eta}}\log g(\boldsymbol{\eta}) \tag{D.7}$$

Comparing (D.7) and (D.4) results

$$\mathbb{E}_q[\mathbf{u}(x)] = \mathbb{E}_p[\mathbf{u}(x)] \tag{D.8}$$

Therefore, the optimum solution to (D.5) satisfies (D.8). If $q(x)$ is selected as a Gaussian distribution, then x and x^2 will be its sufficient statistics. In this case, (D.8) imply the mean and the variance under $q(x)$ should be equal to mean and variance under $p(x)$, respectively. Threfore, this property is known as *moment matching*.

[1] Although continues distributions are considered, similar results hold for discrete distributions.

Appendix E
ADF Equations for Gaussian Random Vectors

Consider the following intermediate PDF for the latent vector $\mathbf{x}$:

$$\hat{p}(\mathbf{x}) = \frac{1}{Z_i} f_i(\mathbf{x}) q^{\setminus i}(\mathbf{x}) \tag{E.1}$$

in which $q^{\setminus i}(\mathbf{x}) \in \mathcal{F}$, and $Z_i = \int_{\mathbf{x}} f_i(\mathbf{x}) q^{\setminus i}(\mathbf{x}) \mathrm{d}\mathbf{x} = \mathbb{E}_{q^{\setminus i}}[f_i(\mathbf{x})]$.

As discussed in ADF algorithm in Appendix C, we are interested in finding a PDF as $q(\mathbf{x})$ which is the best approximation of $\hat{p}(\mathbf{x})$ in $\mathcal{F}$, or in the other word, is its projection in $\mathcal{F}$. Therefore, we must deal with the following problem:

$$q^*(\mathbf{x}) = \underset{q(\mathbf{x}) \in \mathcal{F}}{\arg\min}\, \mathbb{KL}\Big(\hat{p}(\mathbf{x}) || q(\mathbf{x})\Big) \tag{E.2}$$

In this section we solve this optimization problem for the specific case of Gaussian variables.

Suppose $q^{\setminus i}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{m}^{\setminus i}, V^{\setminus i})$ and $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{m}, V)$ are Gaussian distributions. Finding $q(\mathbf{x})$ by moment matching in (D.8) is equivalent to calculate its mean vector and covariance matrix as follows:

$$\mathbf{m} = \mathbb{E}_q[\mathbf{x}] = \mathbb{E}_{\hat{p}}[\mathbf{x}] \tag{E.3}$$

and

$$\begin{aligned} V &= \mathbb{E}_q[\mathbf{x}\mathbf{x}^T] - \mathbb{E}_q[\mathbf{x}]\mathbb{E}_q[\mathbf{x}]^T \\ &= \mathbb{E}_{\hat{p}}[\mathbf{x}\mathbf{x}^T] - \mathbb{E}_{\hat{p}}[\mathbf{x}]\mathbb{E}_{\hat{p}}[\mathbf{x}]^T \end{aligned} \tag{E.4}$$

Therefor we must calculate $\mathbb{E}_{\hat{p}}[\mathbf{x}]$ and $\mathbb{E}_{\hat{p}}[\mathbf{x}\mathbf{x}^T]$. We start with definition of Z_i which is given as

$$Z_i = \frac{1}{|2\pi V^{\setminus i}|^{1/2}} \int_{\mathbf{x}} f_i(\mathbf{x}) \exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big)\mathrm{d}\mathbf{x} \tag{E.5}$$

in which, $\boldsymbol{\omega}_i = \mathbf{x} - \mathbf{m}^{\setminus i}$.

In following parts we calculate two multidimensional gradient $\nabla_m \triangleq \nabla_{\mathbf{m}^{\setminus i}} \log Z_i$ and $\nabla_V \triangleq \nabla_{V^{\setminus i}} \log Z_i$.

Calculating ∇_m:

The differential of $\log Z_i$ with respect to $\mathbf{m}^{\setminus i}$ can be calculated as follows:

$$\begin{aligned}
&\mathrm{d}\log Z_i = \frac{1}{Z_i}\mathrm{d}Z_i \\
&= \frac{1}{Z_i}\frac{1}{|2\pi V^{\setminus i}|^{1/2}} \int_{\mathbf{x}} f_i(\mathbf{x})\mathrm{d}\Big[\exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big)]\mathrm{d}\mathbf{x} \\
&= \frac{1}{Z_i}\frac{1}{|2\pi V^{\setminus i}|^{1/2}} \int_{\mathbf{x}} f_i(\mathbf{x}) \exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big)\mathrm{d}\Big[-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big]\mathrm{d}\mathbf{x} \\
&= \frac{1}{Z_i}\frac{1}{|2\pi V^{\setminus i}|^{1/2}} \int_{\mathbf{x}} f_i(\mathbf{x}) \exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big)\Big[-\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\mathrm{d}\boldsymbol{\omega}_i\Big]\mathrm{d}\mathbf{x} \\
&= \frac{1}{Z_i}\frac{1}{|2\pi V^{\setminus i}|^{1/2}} \int_{\mathbf{x}} f_i(\mathbf{x}) \exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big)\Big[\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\mathrm{d}\mathbf{m}^{\setminus i}\Big]\mathrm{d}\mathbf{x} \\
&= \left(\frac{1}{Z_i}\frac{1}{|2\pi V^{\setminus i}|^{1/2}} \int_{\mathbf{x}} f_i(\mathbf{x}) \exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\boldsymbol{\omega}_i\Big)\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\mathrm{d}\mathbf{x}\right)\mathrm{d}\mathbf{m}^{\setminus i} \\
&= \left(\int_{\mathbf{x}} \hat{p}(\mathbf{x})\boldsymbol{\omega}_i^T (V^{\setminus i})^{-1}\mathrm{d}\mathbf{x}\right)\mathrm{d}\mathbf{m}^{\setminus i} \\
&= \left((V^{\setminus i})^{-1}(\mathbb{E}_{\hat{p}}[\mathbf{x}] - \mathbf{m}^{\setminus i})\right)^T \mathrm{d}\mathbf{m}^{\setminus i}
\end{aligned} \tag{E.6}$$

Therefore,

$$\frac{\partial \log Z_i}{\partial \mathbf{m}^{\setminus i}} = \left((V^{\setminus i})^{-1}(\mathbb{E}_{\hat{p}}[\mathbf{x}] - \mathbf{m}^{\setminus i})\right)^T \tag{E.7}$$

By considering $\nabla_{\mathbf{m}} = \nabla_{\mathbf{m}^{\setminus i}} \log Z_i = (\frac{\partial \log Z_i}{\partial \mathbf{m}^{\setminus i}})^T$, we can write

$$\nabla_m = (V^{\setminus i})^{-1}(\mathbb{E}_{\hat{p}}[\mathbf{x}] - \mathbf{m}^{\setminus i}) \tag{E.8}$$

Finnaly, by (E.3) and (E.8),

$$\mathbf{m} = \mathbb{E}_{\hat{p}}[\mathbf{x}] = \mathbf{m}^{\backslash i} + V^{\backslash i}\nabla_m \tag{E.9}$$

□

Calculating ∇_V:

The differential of $\log Z_i$ with respect to $V^{\backslash i}$ can be calculated as follows:

$$\begin{aligned}
&\mathrm{d}\log Z_i = \mathrm{d}\bigg(\log\int_{\mathbf{x}} f_i(\mathbf{x})\exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\boldsymbol{\omega}_i\Big)\mathrm{d}\mathbf{x} - \frac{1}{2}\log|2\pi V^{\backslash i}|\bigg)\\
&= \frac{1}{Z_i|2\pi V^{\backslash i}|^{1/2}}\mathrm{d}\bigg(\int_{\mathbf{x}} f_i(\mathbf{x})\exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\boldsymbol{\omega}_i\Big)\mathrm{d}\mathbf{x}\bigg)\\
&-\frac{1}{2}\mathrm{tr}\Big((V^{\backslash i})^{-1}\mathrm{d}V^{\backslash i}\Big)\\
&= \frac{1}{Z_i|2\pi V^{\backslash i}|^{1/2}}\times\\
&\int_{\mathbf{x}} f_i(\mathbf{x})\exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\boldsymbol{\omega}_i\Big)(\frac{1}{2})\mathrm{tr}\Big((V^{\backslash i})^{-1}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\mathrm{d}V^{\backslash i}\Big)\mathrm{d}\mathbf{x}\\
&-\frac{1}{2}\mathrm{tr}\Big((V^{\backslash i})^{-1}\mathrm{d}V^{\backslash i}\Big)
\end{aligned} \tag{E.10}$$

Therefore,

$$\begin{aligned}
\frac{\partial\log Z_i}{\partial V^{\backslash i}} &= \frac{1}{2}\frac{1}{Z_i|2\pi V^{\backslash i}|^{1/2}}\int_{\mathbf{x}} f_i(\mathbf{x})\exp\Big(-\frac{1}{2}\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\boldsymbol{\omega}_i\Big)(V^{\backslash i})^{-1}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\mathrm{d}\mathbf{x}\\
&\quad -\frac{1}{2}(V^{\backslash i})^{-1}\\
&= \frac{1}{2}\int_{\mathbf{x}}\hat{p}(\mathbf{x})(V^{\backslash i})^{-1}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T(V^{\backslash i})^{-1}\mathrm{d}\mathbf{x} - \frac{1}{2}(V^{\backslash i})^{-1}\\
&= \frac{1}{2}(V^{\backslash i})^{-1}\bigg(\int_{\mathbf{x}}\hat{p}(\mathbf{x})(\mathbf{x}-\mathbf{m}^{\backslash i})(\mathbf{x}-\mathbf{m}^{\backslash i})^T\mathrm{d}\mathbf{x}\bigg)(V^{\backslash i})^{-1} - \frac{1}{2}(V^{\backslash i})^{-1}\\
&= \frac{1}{2}(V^{\backslash i})^{-1}\bigg(\mathbb{E}_{\hat{p}}[\mathbf{x}\mathbf{x}^T] - \mathbb{E}_{\hat{p}}[\mathbf{x}]\mathbf{m}^{\backslash i^T} - \mathbf{m}^{\backslash i}\mathbb{E}_{\hat{p}}[\mathbf{x}^T] + \mathbf{m}^{\backslash i}\mathbf{m}^{\backslash i^T}\bigg)(V^{\backslash i})^{-1}\\
&\quad -\frac{1}{2}(V^{\backslash i})^{-1}
\end{aligned} \tag{E.11}$$

By considering $\nabla_V = \nabla_{V^{\setminus i}} \log Z_i = (\frac{\partial \log Z_i}{\partial V^{\setminus i}})^T$, we can write

$$\mathbb{E}_{\hat{p}}[\mathbf{x}\mathbf{x}^T] = 2V^{\setminus i}\nabla_V V^{\setminus i} + \mathbb{E}_{\hat{p}}[\mathbf{x}]\mathbf{m}^{\setminus i^T} + \mathbf{m}^{\setminus i}\mathbb{E}_{\hat{p}}[\mathbf{x}^T] - \mathbf{m}^{\setminus i}\mathbf{m}^{\setminus i^T} + V^{\setminus i} \tag{E.12}$$

By using (E.9) and plugging the equivalent expression of $\mathbb{E}_{\hat{p}}[\mathbf{x}]$, we can write

$$\begin{aligned} \mathbb{E}_{\hat{p}}[\mathbf{x}\mathbf{x}^T] &= 2V^{\setminus i}\nabla_V V^{\setminus i} + (\mathbf{m}^{\setminus i} + V^{\setminus i}\nabla_m)\mathbf{m}^{\setminus i^T} + \mathbf{m}^{\setminus i}(\mathbf{m}^{\setminus i} + V^{\setminus i}\nabla_m)^T \\ &\quad -\mathbf{m}^{\setminus i}\mathbf{m}^{\setminus i^T} + V^{\setminus i} \end{aligned} \tag{E.13}$$

By (E.4),

$$\begin{aligned} V &= 2V^{\setminus i}\nabla_V V^{\setminus i} + (\mathbf{m}^{\setminus i} + V^{\setminus i}\nabla_m)\mathbf{m}^{\setminus i^T} + \mathbf{m}^{\setminus i}(\mathbf{m}^{\setminus i} + V^{\setminus i}\nabla_m)^T \\ &\quad -\mathbf{m}^{\setminus i}\mathbf{m}^{\setminus i^T} + V^{\setminus i} - (\mathbf{m}^{\setminus i} + V^{\setminus i}\nabla_m)(\mathbf{m}^{\setminus i} + V^{\setminus i}\nabla_m)^T \\ &= 2V^{\setminus i}\nabla_V V^{\setminus i} - V^{\setminus i}\nabla_m\nabla_m^T V^{\setminus i} + V^{\setminus i} \end{aligned} \tag{E.14}$$

which can finaly expressed as

$$V = V^{\setminus i} - V^{\setminus i}(\nabla_m\nabla_m^T - 2\nabla_V)V^{\setminus i} \tag{E.15}$$

□

Equations (E.9) and (E.15) are known as ADF equations and are very useful in updating statistics of multidimensional Gaussian random variables in ADF or expectation propagation algorithms.

Appendix F
Calculating Z_t

The constant Z_t can be calculated as follows:

$$
\begin{aligned}
Z_t &= \mathbb{E}_{q^{\setminus t}(\tilde{S},\mathbf{h})}[p(\mathbf{y}_t|\tilde{\mathbf{s}}_t,\mathbf{h})] \\
&= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} \int q^{\setminus t}(\tilde{S},\mathbf{h})p(\mathbf{y}_t|\tilde{\mathbf{s}}_t,\mathbf{h})\mathrm{d}\mathbf{h} \\
&= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} \int q^{\setminus t}(\tilde{S},\mathbf{h})p(\mathbf{y}_t|\tilde{\mathbf{s}}_t,\mathbf{h})\mathrm{d}\mathbf{h} \\
&= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} q^{\setminus t}(\tilde{S}) \int \mathcal{N}\Big(\mathbf{h}|\mathbf{m}_h^{\setminus t},V_h^{\setminus t}\Big)\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{h},\frac{1}{2}\sigma_w'^2 I_{2M}\Big)\mathrm{d}\mathbf{h} \\
&= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} q^{\setminus t}(\tilde{S})\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t},S_tV_h^{\setminus t}S_t^T+\frac{1}{2}\sigma_w'^2 I_{2M}\Big) \\
&= \sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K} \left(\prod_{\substack{i=1\\i\neq t}}^{T} q_i(\tilde{\mathbf{s}}_i)\right)\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t},S_tV_h^{\setminus t}S_t^T+\frac{1}{2}\sigma_w'^2 I_{2M}\Big) \\
&= \left(\sum_{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_{t-1},\tilde{\mathbf{s}}_{t+1},\ldots,\tilde{\mathbf{s}}_T\in\tilde{\mathcal{A}}_{\mathcal{M}}^K}\prod_{\substack{i=1\\i\neq t}}^{T} q_i(\tilde{\mathbf{s}}_i)\right)\sum_{\tilde{\mathbf{s}}_t\in\tilde{\mathcal{A}}_M^K}\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t},S_tV_h^{\setminus t}S_t^T+\frac{1}{2}\sigma_w'^2 I_{2M}\Big) \\
&= \sum_{\tilde{\mathbf{s}}_t\in\tilde{\mathcal{A}}_M^K}\mathcal{N}\Big(\mathbf{y}_t|S_t\mathbf{m}_h^{\setminus t},S_tV_h^{\setminus t}S_t^T+\frac{1}{2}\sigma_w'^2 I_{2M}\Big) \qquad \text{(F.1)}
\end{aligned}
$$

Appendix G
Hybrid K-L Divergence Optimization Problem

Consider the following optimization problem on the hybrid K-L divergence:

$$\min_{q(S,\mathbf{h})\in\ \mathcal{F}} \mathrm{KL}\Big(\hat{p}_t(S,\mathbf{h})||q(S)q(\mathbf{h})\Big), \tag{G.1}$$

in which, $S=[\mathbf{s}_1,\ldots,\mathbf{s}_T]$ is a $K\times T$ matrix of discrete symbols in $\mathcal{A}_{\mathcal{M}}$, i.e. $S\in\mathcal{A}_{\mathcal{M}}^{K\times T}$, and $\mathbf{h}$ is the continues channel vector. By using the definition of K-L divergence, we can write

$$\begin{aligned}
\mathrm{KL}\Big(\hat{p}_t(S,\mathbf{h})||q(S)q(\mathbf{h})\Big) &= \mathbb{E}_{\hat{p}_t(S,\mathbf{h})}\Big[\log\frac{\hat{p}_t(S,\mathbf{h})}{q(S)q(\mathbf{h})}\Big] \\
&= \sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T\in\mathcal{A}_{\mathcal{M}}^K}\int_{\mathbf{h}}\hat{p}_t(S,\mathbf{h})\log\frac{\hat{p}_t(S,\mathbf{h})}{q(S)q(\mathbf{h})}\mathrm{d}\mathbf{h} \\
&= \sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T\in\mathcal{A}_{\mathcal{M}}^K}\int_{\mathbf{h}}\hat{p}_t(S,\mathbf{h})\log\hat{p}_t(S,\mathbf{h})\mathrm{d}\mathbf{h} - \\
&\quad \sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T\in\mathcal{A}_{\mathcal{M}}^K}\int_{\mathbf{h}}\hat{p}_t(S,\mathbf{h})\log\Big(q(S)q(\mathbf{h})\Big)\mathrm{d}\mathbf{h}
\end{aligned} \tag{G.2}$$

Since the optimization in (G.1) is over the functional $q(S,\mathbf{h})$, the first term in (G.2) will be a constant with respect to $q(S,\mathbf{h})$ and can be replaced with letter C. Therefore we can continue (G.2) as follows:

$$\begin{aligned}
\mathrm{KL}\Big(\hat{p}_t(S,\mathbf{h})||q(S)q(\mathbf{h})\Big) &= -\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T\in\mathcal{A}_{\mathcal{M}}^K}\int_{\mathbf{h}}\hat{p}_t(S,\mathbf{h})\log\Big(q(S)q(\mathbf{h})\Big)\mathrm{d}\mathbf{h}+C \\
&= -\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T\in\mathcal{A}_{\mathcal{M}}^K}\int_{\mathbf{h}}\hat{p}_t(S,\mathbf{h})\log q(S)\mathrm{d}\mathbf{h} \\
&\quad -\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T\in\mathcal{A}_{\mathcal{M}}^K}\int_{\mathbf{h}}\hat{p}_t(S,\mathbf{h})\log q(\mathbf{h})\mathrm{d}\mathbf{h}+C
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T \in \mathcal{A}_{\mathcal{M}}^K} \log q(S) \int_{\mathbf{h}} \hat{p}_t(S,\mathbf{h}) \mathrm{d}\mathbf{h} \\
&\quad - \int_{\mathbf{h}} \log q(\mathbf{h}) \Big(\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T \in \mathcal{A}_{\mathcal{M}}^K} \hat{p}_t(S,\mathbf{h}) \Big) \mathrm{d}\mathbf{h} + C \\
&= -\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T \in \mathcal{A}_{\mathcal{M}}^K} \hat{p}_t(S) \log q(S) \\
&\quad - \int_{\mathbf{h}} \hat{p}_t(\mathbf{h}) \log q(\mathbf{h}) \mathrm{d}\mathbf{h} + C
\end{aligned} \tag{G.3}
$$

in which,

$$
\hat{p}_t(S) = \int_{\mathbf{h}} \hat{p}_t(S,\mathbf{h}) \mathrm{d}\mathbf{h} \tag{G.4}
$$

and

$$
\hat{p}_t(\mathbf{h}) = \sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T \in \mathcal{A}_{\mathcal{M}}^K} \hat{p}_t(S,\mathbf{h}) \tag{G.5}
$$

are respective marginal distributions of S and $\mathbf{h}$ from $\hat{p}_t(S,\mathbf{h})$.

By adding and subtracting two constant terms $\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T \in \mathcal{A}_{\mathcal{M}}^K} \hat{p}_t(S) \log \hat{p}_t(S)$ and $\int_{\mathbf{h}} \hat{p}_t(\mathbf{h}) \log \hat{p}_t(\mathbf{h}) \mathrm{d}\mathbf{h}$ to (G.3), we can write it as

$$
\begin{aligned}
\mathrm{KL}\Big(\hat{p}_t(S,\mathbf{h}) || q(S)q(\mathbf{h})\Big) &= \sum_{\mathbf{s}_1,\ldots,\mathbf{s}_T \in \mathcal{A}_{\mathcal{M}}^K} \hat{p}_t(S) \log \frac{\hat{p}_t(S)}{q(S)} + \\
&\quad \int_{\mathbf{h}} \hat{p}_t(\mathbf{h}) \log \frac{\hat{p}_t(\mathbf{h})}{q(\mathbf{h})} \mathrm{d}\mathbf{h} + C' \\
&= \mathrm{KL}\Big(\hat{p}_t(S) || q(S)\Big) + \mathrm{KL}\Big(\hat{p}_t(\mathbf{h}) || q(\mathbf{h})\Big) + C'
\end{aligned} \tag{G.6}
$$

The first term in (G.6) totally depends on $q(\mathbf{s})$ and the second term only depends on $q(S)$. Therefore the first two terms of (G.3) are changing independently. Therefore, the minimization of (G.3) will be equivalent to minimization of both of these two terms. Consequently, (G.1) can be broken into the following optimization problems:

$$
q^{new}(S) = \arg\min_{q(S)} \mathrm{KL}\Big(\hat{p}_t(S) || q(S)\Big) \tag{G.7}
$$

and

$$q^{new}(\mathbf{h}) = \arg\min_{q(\mathbf{h})} \mathrm{KL}\Big(\hat{p}_t(\mathbf{h})||q(\mathbf{h})\Big) \tag{G.8}$$

Vita

Kamran Ghavami received his B.S. and M.S. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1998 and 2000, respectively. Upon the completion of the M.S. degree, he conducted research and implementation of telecommunication systems as a Research Engineer at NRI, where he participated in several related projects. In January 2013, he began working toward the Ph.D. degree with Louisiana State University, Baton Rouge, LA, USA. His research interests include channel coding, iterative detection, and multiple-input multiple-output wireless systems.