

TECHNICAL UNIVERSITY OF CRETE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
TELECOMMUNICATIONS DIVISION



Approximate Message Passing for Joint Data and Channel Estimation in Multi-cell Massive MIMO

by

Nikolaos Psaromanolakis

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE MASTER OF SCIENCE OF
ELECTRICAL AND COMPUTER ENGINEERING

November 2017

THESIS COMMITTEE

Associate Professor Aggelos Bletsas, *Thesis Supervisor*
Professor Michael Paterakis
Associate Professor George N. Karystinos

Abstract

This work studies joint channel estimation and data detection in uplink, multi-cell (interference-limited), large-scale antenna (i.e., massive multiple input multiple output - MIMO) networks. With massive MIMO, future 5G networks will be able to communicate more data within a given spectrum. Realistic assumptions are considered, accounting for path-loss and pilot contamination from adjacent cells.

First, a baseline prior art algorithm is presented, that initially utilizes minimum mean-squared error (MMSE) channel estimator and then linear MMSE detector (given estimated channel) for data detection (abbreviated as SCED). Second, a novel iterative algorithm for joint channel estimation and linear data detection (JCED-Iterative) is proposed, where estimation and detection are done iteratively, until convergence. Third, based on joint data/channel estimation bilinear generalized approximate message passing (JCD-GAMP) proposed recently, which is an approximation of the sum-product belief propagation algorithm and offers scalar computation operations, two inference-based algorithms for joint MMSE channel and data estimation are developed; the first algorithm is an extension of JCD-GAMP to the multi-cell setting, exploiting bilinear approximate message passing and pilot symbols used both in the initialization, as well as the message passing phase (MC-JCD-GAMP); the second inference algorithm uses pilot symbols only for initialization (MC-JCD-GAMP-D).

Simulation results underline the critical importance of damping and initialization fine-tuning in the convergence of the inference-based algorithms. It is found that joint channel estimation and detection algorithms, in few iterations, can achieve better performance compared to SCED. Moreover, it is shown that pilot symbols are helpful in the initialization phase and not in the subsequent message passing procedure, i.e. MC-JCD-GAMP and MC-

JCD-GAMP-D offer similar results. The two latter have better performance compared to JCED-Iterative at high SNR, but JCED-Iterative algorithm converges faster in all cases. Hence, there exist interesting convergence-accuracy and tuning-complexity trade-offs between the above practical algorithms for massive MIMO uplink networks.

Acknowledgements

First of all I would like to thank my family for their endless support and love throughout my years of study. They have always been there for me and without them this thesis would have never been completed.

I would like to thank my supervisor Prof. Aggelos Bletsas for his encouragement and his remarks during all the years of our cooperation. He has always been willing to provide the best and to the point advices, and he gave me the opportunity to be part of an extraordinary group.

Furthermore, I would like to express my gratitude to all the group members I have collaborated with, and especially Panos Alevizos who dedicated a lot of his time to willingly offer his help and also to provide me with his knowledge.

Finally, many thanks to all my friends for their support and for the great moments that we have lived together.

To my family.

Table of Contents

Table of Contents	6
List of Figures	8
1 Introduction	11
1.1 5G and Massive MIMO	11
1.2 Channel Estimation and Data Detection	11
1.3 Thesis Contribution	13
1.4 Thesis Outline	14
2 Problem Statement and System Model	15
2.1 System Model	15
2.1.1 Path-Loss Model and Rayleigh Fading	15
2.1.2 Pilot and Data Symbols	18
2.2 Problem Formulation - Joint Data and Channel Estimation	19
3 Closed-form Channel Estimation and Linear Data Detection	21
3.1 Separate MMSE Estimation and Linear MMSE Detection	21
3.1.1 Channel Estimation	21
3.1.2 Linear Detection	22
3.1.3 Computation Complexity	23
3.2 Iterative Estimation and Linear Detection	24
3.2.1 Channel Estimation	25
3.2.2 Linear Detection	26
3.2.3 Iterative Procedure	26
3.2.4 Computation Complexity	26

4 Bilinear Generalized Approximate Message Passing for MMSE	
Joint Data and Channel Estimation	30
4.1 Factor Graph	31
4.2 MC-JCD-GAMP Algorithm	32
4.2.1 Initialization	35
4.2.2 Estimators	35
4.2.3 Damping Factor	39
4.2.4 Computation Complexity	40
4.3 MC-JCD-GAMP-D Algorithm	40
4.3.1 Computation Complexity	42
5 Simulation Results	45
5.1 Impact of Damping Factor	46
5.2 First Scenario (64 or 128 Received Antennas, 10 UTs per Cell)	48
5.3 Second Scenario (64 Received Antennas, 15 UTs per Cell)	55
5.4 Third Scenario (Moving Intra-cell UT)	58
6 Conclusions and Future Work	62
6.1 Conclusions	62
6.2 Future Work	63

Appendices

A Mean and Variance of Noise in JCED-Iterative	65
B Mean and Variance of Interference Signal	67
C Product of Two Gaussian Distributions	70
D Estimator for QPSK Data Symbols	72
Bibliography	75

List of Figures

2.1	Mutlicell network of $L = 7$ cells, frequency reuse factor = 1, $K = 2$ UTs per cell, with representation of useful and interference links at central cell.	16
2.2	Mutlicell network of $L = 19$ cells, frequency reuse factor = 3, $K = 2$ UTs per cell, with representation of useful and interference links at central cell.	17
3.1	Flow diagram of JCED-Iterative algorithm.	29
4.1	Graphical illustrator of factor graph for MC-JCD-GAMP algorithm, where $N = 2$, $K = 2$, $T_t = 1$ and $T_d = 1$	32
4.2	Representation of MC-JCD-GAMP algorithm.	35
4.3	Flow diagram of MC-JCD-GAMP-D algorithm.	41
4.4	Graphical illustrator of factor graph for the MC-JCD-GAMP-D algorithm, where $N = 2$, $K = 2$ and $T_d = 1$	43
5.1	Topology of 7-cell hexagonal system with $K = 10$ UTs uniformly distributed around each BS.	46
5.2	Divergence outage probability versus damping factor, for $N = 64$, $K = 10$ and $T_d = 240$	47
5.3	Iterations until convergence versus damping factor, for $N = 64$, $K = 10$ and $T_d = 240$	48
5.4	Symbol Error Rate versus average received SNR, for $N = 64$, $K = 10$ and $T_d = 240$	49
5.5	Symbol Error Rate versus minimum received SNR, for $N = 64$, $K = 10$ and $T_d = 240$	50

5.6	Symbol Error Rate versus maximum received SNR, for $N = 64$, $K = 10$ and $T_d = 240$	50
5.7	Symbol Error Rate versus transmit power per intra-cell UT, for $N = 64$, $K = 10$ and $T_d = 240$	51
5.8	NRMSE versus average received SNR, for $N = 64$, $K = 10$ and $T_d = 240$	51
5.9	Iterations until convergence versus average received SNR, for $N = 64$, $K = 10$ and $T_d = 240$	52
5.10	Symbol Error Rate versus average received SNR, for $N = 64$ or $N = 128$, $K = 10$ and $T_d = 240$	53
5.11	NRMSE versus average received SNR, for $N = 64$ or $N = 128$, $K = 10$ and $T_d = 240$	53
5.12	Iterations until convergence versus average received SNR, for $N = 64$ or $N = 128$, $K = 10$ and $T_d = 240$	54
5.13	Symbol Error Rate versus average received SNR, for $N = 64$, $K = 15$ and $T_d = 360$	55
5.14	Symbol Error Rate versus transmit power per intra-cell UT, for $N = 64$, $K = 15$ and $T_d = 360$	56
5.15	NRMSE versus average received SNR, for $N = 64$, $K = 15$ and $T_d = 360$	57
5.16	Iterations until convergence versus average received SNR, for $N = 64$, $K = 15$ and $T_d = 360$	57
5.17	Topology of 7-cell hexagonal system with $K = 15$ UTs uniformly distributed around each BS, where arrow represents the movement of intra-cell UT k	58
5.18	Symbol Error Rate versus received SNR from intra-cell UT k , for $N = 64$, $K = 15$ and $T_d = 360$	59
5.19	Symbol Error Rate versus distance of intra-cell UT k from BS, for $N = 64$, $K = 15$ and $T_d = 360$	60
5.20	NRMSE versus received SNR from intra-cell UT k , for $N = 64$, $K = 15$ and $T_d = 360$	60
5.21	Iterations until convergence versus received SNR from intra-cell UT k , for $N = 64$, $K = 15$ and $T_d = 360$	61

Notations

x	variable
\mathbf{x}	a vector
\mathbf{A}	a matrix
\mathbf{A}^T	transpose of matrix \mathbf{A}
\mathbf{A}^H	the conjugate transpose (hermitian) of matrix \mathbf{A}
x^*	the conjugate of variable x
$\ \mathbf{x}\ _p$	the p norm of vector \mathbf{x}
$\ \mathbf{x}\ _F$	the Frobenius norm of vector \mathbf{x}
\mathbf{I}_n	the $n \times n$ identity matrix
\mathbb{C}	the set of complex numbers
$\mathcal{N}_{\mathbb{C}}(x; \mu, \Sigma)$	the complex random variable x drawn from the proper Gaussian distribution of mean μ and variance Σ
$\mathcal{N}_{\mathbb{C}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	the complex random vector \mathbf{x} drawn from the proper Gaussian distribution of mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$
$f_{y x}(y x)$	a conditional pdf of random variable y given the variable x
$f_{\mathbf{y} \mathbf{x}}(\mathbf{y} \mathbf{x})$	a conditional pdf of random vector \mathbf{y} given the vector \mathbf{x}

Chapter 1

Introduction

1.1 5G and Massive MIMO

Fifth-generation (5G) cellular networks target to offer larger capacity (1000 times increase) than current 4G, allowing a higher density of mobile broadband users [1, 2]. Therefore, in order to accomplish the above task, one possible solution is the use of large-scale multiple-input multiple-output antenna systems. In massive MIMO systems, base stations (BSs) are equipped with hundred or thousands of antennas, hence they can serve hundred of user terminals (UTs) that share the same frequency resources [3–6]. Furthermore, due to the fact that the number of antennas at the BS is much larger compared to single antenna UTs, the use of the simplest linear detectors and precoders is optimal. Therefore, interference caused by adjacent cells due to reuse of pilot sequences is the only performance limitation, while interference, channel estimation errors and thermal noise are almost vanished [7].

1.2 Channel Estimation and Data Detection

The current work focuses on the channel estimation and data detection problem for multi-cell uplink massive MIMO cellular networks. The two classic ways of channel estimation and data detection are with the joint and with the separate procedure. In [8, 9] estimation schemes based on suboptimal criteria for joint channel estimation and data detection are proposed. It is described that with joint channel estimation and data detection, only a few pilot symbols are needed in order to achieve estimation and detection in a satisfying level. In [10], an iterative genetic algorithm that provide soft outputs is proposed for joint channel estimation and detection in MIMO net-

work. Unlike conventional methods where pilots and data symbols separate into two transmit phases, a pilot embedding method where low-level pilots and data are transmitted together in order to obtain an initial estimate of the channel is proposed in [11]. Authors in [7] describe an algorithm that follows the separate procedure for channel estimation from data detection under realistic assumption for the system model, where path-loss and pilot contamination from adjacent cells are taken into account. That algorithm is further analyzed in Sec. 3.1.

In [12], a Bayes-optimal inference algorithm for joint channel estimation and data detection in single cell uplink massive MIMO networks with low-precision ADCs is proposed. The above algorithm is based on Bilinear Generalized Approximate Message Passing (Bilinear-GAMP) algorithm, which uses scalar operations with simple computational methods in order to compute the estimated values [13]. Bilinear-GAMP algorithm is defined as an approximation of the sum-product belief propagation algorithm that operates in a factor graph [14] in the high-dimensional limit, where Taylor-series approximations and central-limit theorem arguments are used. It is shown that Bayes-optimal inference algorithm can generally describe the performance of the theoretical Bayes-optimal estimator. Hence, Bayes-optimal inference algorithm exhibits the best possible performance.

In this work, schemes for joint channel estimation and data detection in multi-cell uplink massive MIMO are provided. Heuristic iterative joint channel estimation and data detection algorithm based on [3] is proposed, while complexity analysis is supplied. The above algorithm uses the pilot symbols along with the detected data symbols in order to estimate, through an iterative procedure, the instantaneous values of the channels between the BS and the intra-cell user terminals (UT), as well as to detect again the data symbols. Furthermore, the Bayes-optimal inference algorithm [12], which uses the pilot symbols during the message passing procedure, is adjusted to multi-cell massive MIMO network. In addition, based on the above algorithm, a new Bayes-optimal inference algorithm for joint channel estimation and data detection is proposed, where the pilot symbols are used only during the initialization phase and the message passing procedure, applied only in

data phase, i.e., the time period during user terminals send their data symbols. Analysis for the damping factor in Bayes-optimal inference algorithms (or BiGAMP-based algorithms) is provided, while computation complexity is also examined.

1.3 Thesis Contribution

The contribution of this thesis is summarized in the following bullets:

- Algorithm that first uses MMSE estimator for channel estimation and then linear MMSE detector for data detection (SCED) is studied [7].
- An heuristic iterative algorithm, based on SCED algorithm, for joint channel estimation and linear data detection (JCED-Iterative) is proposed. Detected data symbols both with pilot symbols are used for the channel estimation.
- Two inference BiGAMP-based [13] algorithms are proposed:
 - JCD-GAMP algorithm [12] is modified (MC-JCD-GAMP) for multi-cell interference limited massive MIMO network and new initialization is proposed. The above algorithm uses pilot symbols both in initialization and message passing phase.
 - MC-JCD-GAMP-D algorithm that uses pilot symbols only in initialization phase and message passing is proceed using the data symbols is proposed. Hence, algorithm has smaller computation complexity in message passing phase compared to MC-JCD-GAMP algorithm.
- Computation complexity analysis is provided for all algorithms (SCED, JCED-Iterative, MC-JCD-GAMP, MC-JCD-GAMP-D).
- Computer simulations that illustrate the impact of damping factor parameter on the convergence in BiGAMP-based algorithms are presented.

-
- Computer simulations under different scenarios with realistic assumptions and system dimensions are performed to verify the efficiency of the above algorithms.
 - Several insightful observations are obtained and numerous of interesting convergence-accuracy and tuning-complexity trade-offs between the studied algorithms are determined.

1.4 Thesis Outline

The thesis is organized as follows: Chapter 2 describes the system model and the problem formulation for multi-cell massive MIMO network. Chapter 3 presents both the separate and the heuristic iterative channel estimation and linear data detection (SCED and JCED-Iterative) algorithms. Chapter 4 presents the bilinear generalized approximate message passing for joint MMSE channel and data estimation (MC-JCD-GAMP and MC-JCD-GAMP-D) algorithms. Chapter 5 offers simulation results and observations that have emerged. Thesis is concluded at Chapter 6.1, in which ideas for future work are also provided.

Chapter 2

Problem Statement and System Model

2.1 System Model

A hexagonal multicellular massive MIMO uplink system with $L > 1$ cells is assumed, i.e. Fig. 2.1, Fig. 2.2¹. The BSs are equipped with N antennas and K UTs single-antenna users. We assume flat block fading channels, which remain constant over T consecutive symbols intervals. All K users are assumed perfectly synchronised. The received signal at the BS j $\mathbf{Y}_j \in \mathbb{C}^{N \times T}$ based on [7] can be expressed as

$$\mathbf{Y}_j = \mathbf{H}_{jj}\mathbf{X}_j + \sum_{l \neq j} \mathbf{H}_{jl}\mathbf{X}_l + \mathbf{W}_j \quad (2.1)$$

where $\mathbf{H}_{jl} = [\mathbf{h}_{jl,1} \ \mathbf{h}_{jl,2} \ \dots \ \mathbf{h}_{jl,K}] \in \mathbb{C}^{N \times K}$, $\mathbf{h}_{jl,k} \in \mathbb{C}^N$ is the channel from UT k in cell l to BS j , $\mathbf{X}_l = [\mathbf{x}_{l,1} \ \mathbf{x}_{l,2} \ \dots \ \mathbf{x}_{l,K}]^T \in \mathbb{C}^{K \times T}$ with $\mathbf{x}_{l,k} \in \mathbb{C}^T$ the transmit symbols of k UT in cell l and $\mathbf{W}_j \in \mathbb{C}^{N \times T}$ is the additive white Gaussian noise at receiver at BS j , where for each element of \mathbf{W}_j is considered that $\mathcal{N}_{\mathbb{C}}(W_j; 0, \sigma_w^2)$. In addition, $\mathbf{Z}_j = \mathbf{H}_{jj}\mathbf{X}_j \in \mathbb{C}^{N \times T}$.

2.1.1 Path-Loss Model and Rayleigh Fading

It is difficult to obtain a simple model to characterize the pass-loss across different environments due to complexity of signal propagation. However, for various systems designs, a simple model that represents the nature of

¹Frequency reuse factor is the rate at which the same frequency can be used by the cells in the network. Hence interference is emanated only from cells that use the same frequency.

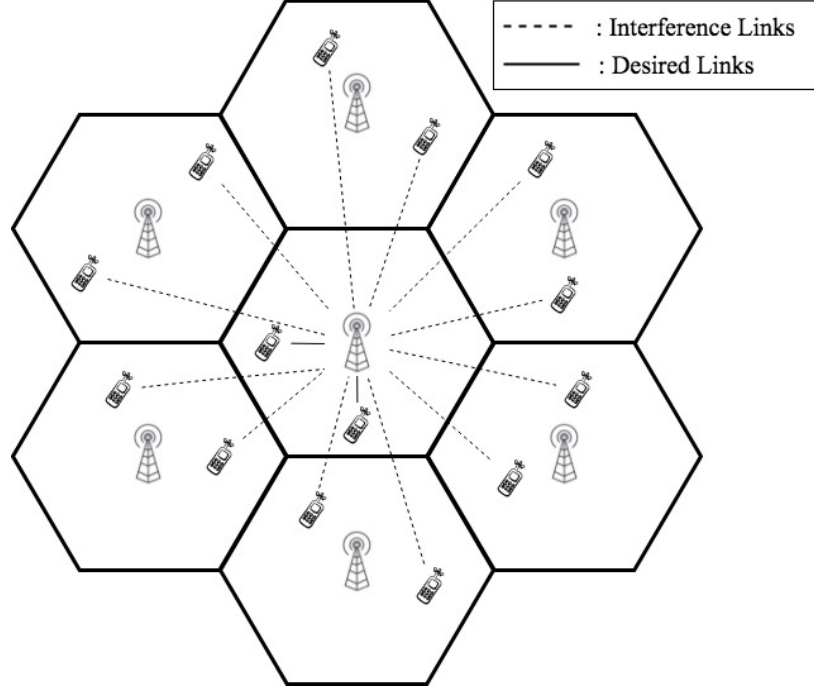


Figure 2.1: Mutlicell network of $L = 7$ cells, frequency reuse factor = 1, $K = 2$ UTs per cell, with representation of useful and interference links at central cell.

signal propagation without including complicated path-loss models is used. Therefore, the following simplified path-loss model as a function of distance according to [15] is defined as

$$P_L = G_{\text{BS}} G_{\text{UT}} \left(\frac{\lambda}{4\pi d_0} \right)^2 \left(\frac{d_0}{d} \right)^\alpha \quad (2.2)$$

where P_L , G_{BS} and G_{UT} are the inverse free space path-loss, the antenna gains at BS and UTs, respectively. Parameter λ is the free space propagation wavelength that depends on carrier frequency, d_0 is a reference distance, $d > d_0$ is the transmission distance and α is the path-loss exponent.

The entries $h_{jl,nk}$ of channel vectors $\mathbf{h}_{jl,k} \in \mathbb{C}^N$, for $n = 1, \dots, N$, $j, l = 1, \dots, L$ and $k = 1, \dots, K$, are assumed $\mathcal{N}_{\mathbb{C}}(h_{jl,nk}; 0, \sigma_{h_{jl,k}}^2)$ so their magnitude

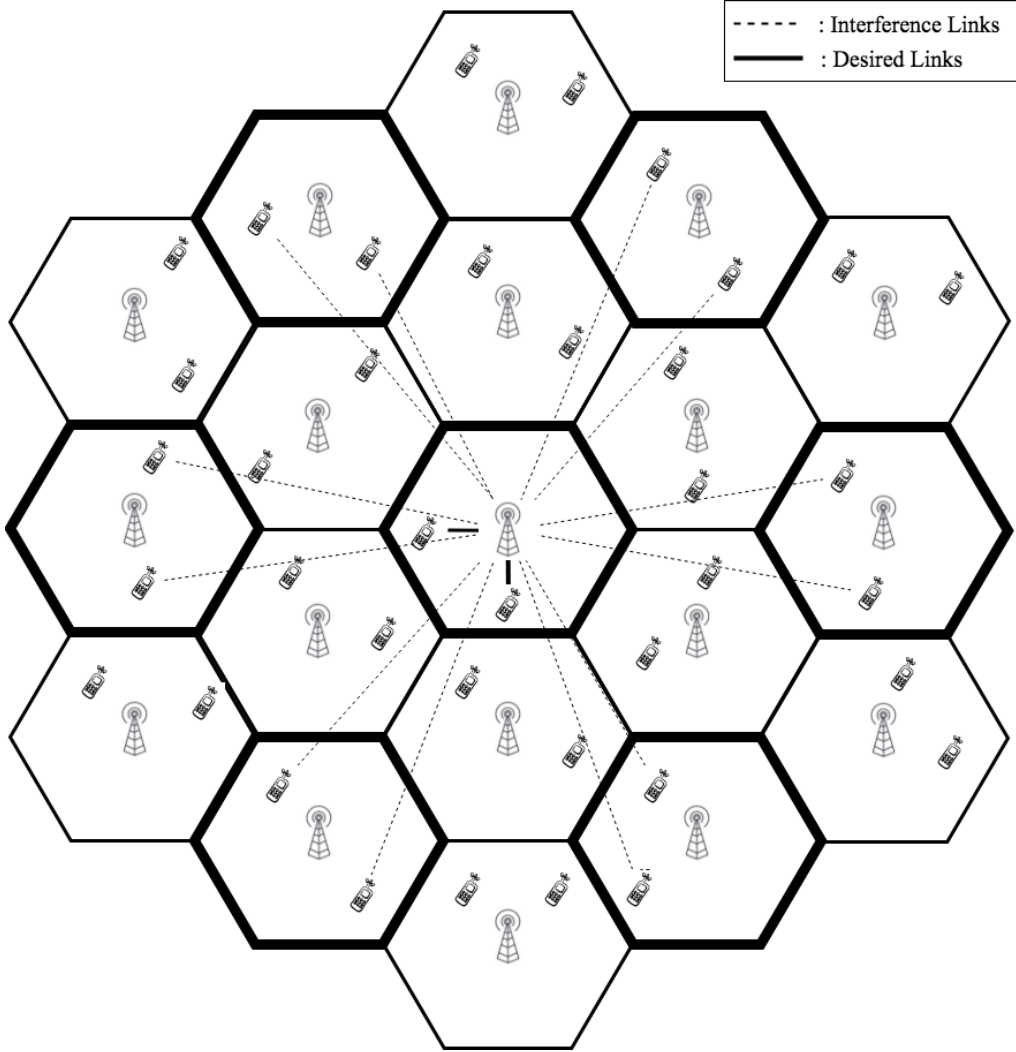


Figure 2.2: Mutlicell network of $L = 19$ cells, frequency reuse factor = 3, $K = 2$ UTs per cell, with representation of useful and interference links at central cell.

follows Rayleigh distribution [15]

$$f_{|h_{jl,nk}|}(x) = \frac{2x}{\sigma_{h_{jl,k}}^2} \exp \left[-\frac{x^2}{\sigma_{h_{jl,k}}^2} \right], x \geq 0 \quad (2.3)$$

where $\sigma_{h_{jl,k}}^2 = \mathbb{E}[|h_{jl,nk}|^2]$ is the average received signal power from UT k of l -th cell to the BS of j -th cell, which takes into account the transmit power

of k UT, the shadowing and the path-loss model that is derived in Eq. (2.2). In more detail, for simplicity in exposition, the channel variance incorporates path-loss shadowing and transmit power. Hence, $\sigma_{h_{jl,k}}^2 = P_{l,k}^{\text{transmit}} P_{L_{jl,k}}$, where $P_{l,k}^{\text{transmit}}$ is the transmit power of k UT in l cell.

Furthermore, the entries of $\mathbf{h}_{jl,k}$ are independent and identically distributed (i.i.d.), thus the prior distribution of \mathbf{H}_{jl} is

$$f_{\mathbf{H}_{jl}}(\mathbf{H}_{jl}) = \prod_{n=1}^N \left(\prod_{k=1}^K \left(f_{h_{jl,nk}}(h_{jl,nk}) \right) \right). \quad (2.4)$$

2.1.2 Pilot and Data Symbols

Symbol transmitting is consisted of two phases: the training phase or t-phase which refers to the first T_t symbols of the T block, serving as pilot sequences, and the data phase or d-phase which refers to $T_d = T - T_t$ remaining symbols. Thus, for $l = 1, \dots, L$, \mathbf{X}_l can be decomposed

$$\mathbf{X}_l = \begin{bmatrix} \mathbf{X}_l^t & \mathbf{X}_l^d \end{bmatrix}, \text{ with } \mathbf{X}_l^t \in \mathbb{C}^{K \times T_t}, \mathbf{X}_l^d \in \mathbb{C}^{K \times T_d}. \quad (2.5)$$

\mathbf{X}_l^t is composed of orthogonal pilot sequences. For their production, Zadoff-Chu (ZC) sequence [16] is used. The analysis of the above procedure is described in [17]. Note that Zadoff-Chu (ZC) sequences are used in 3GPP and LTE standard [18]. Pilot contamination from adjacent cells is assumed, since the same set of orthogonal pilot sequences is reused in every cell. Pilot sequences are deterministic, therefore \mathbf{X}_l^t can be described by probability distribution below

$$f_{\mathbf{X}_l^t}(\mathbf{X}_l^t) = \prod_{k=1}^K \left(\prod_{t=1}^{T_t} \left(f_{X_{l,kt}^t}(X_{l,kt}^t) \right) \right) \quad (2.6)$$

where $f_{X_{l,kt}^t}(X_{l,kt}^t) = \delta(X_{l,kt}^t)$, δ is the Dirac delta function.

\mathbf{X}_l^d is composed of i.i.d. random variables with probability distribution ²

$$f_{\mathbf{X}_l^d}(\mathbf{X}_l^d) = \prod_{k=1}^K \left(\prod_{t=1}^{T_d} \left(f_{X_{l,kt}^d}(X_{l,kt}^d) \right) \right). \quad (2.7)$$

Variables \mathbf{X}_l^d are selected from QPSK modulation. So the average of $X_{l,kt}^d$ are assumed to be zero. However, \mathbf{X}_l^t and \mathbf{X}_l^d are independent and prior distribution for \mathbf{X}_l is given by

$$f_{\mathbf{X}_l}(\mathbf{X}_l) = \prod_{k=1}^K \left(\prod_{t=1}^{T_t} \left(f_{X_{l,kt}^t}(X_{l,kt}^t) \right) \right) \times \prod_{k=1}^K \left(\prod_{t=1}^{T_d} \left(f_{X_{l,kt}^d}(X_{l,kt}^d) \right) \right). \quad (2.8)$$

2.2 Problem Formulation - Joint Data and Channel Estimation

Our goal is to detect $\mathbf{X}_l^d \in \mathbb{C}^{K \times T_d}$, which are the data symbols of each UT in cell j and to estimate $\mathbf{H}_{jj} \in \mathbb{C}^{N \times K}$, which are the instantaneous values of each channel between UT k and BS in cell j . Moreover, the minimum value for the training duration is considered, in order to calculate an acceptable channel state information (CSI). The same band of frequencies is shared to all cells. Each cell serves the same (maximum) number of UT, so the pilot signals received by BS j are contaminated by pilots transmitted by UT in the other cells. If we treat \mathbf{X}_l^d and \mathbf{H}_{jj} as independent random variables with known separable prior distribution as defined in Eq. (2.7), Eq. (2.4), and given that $\mathbf{Z}_j = \mathbf{H}_{jj}\mathbf{X}_j \in \mathbb{C}^{N \times T}$, the distribution of received signal in Eq. (2.1) under unknown parameters is

$$f_{\mathbf{Y}_j|\mathbf{Z}_j}(\mathbf{Y}_j | \mathbf{Z}_j) = \prod_{n=1}^N \left(\prod_{k=1}^K \left(f_{Y_{j,nk}|Z_{j,nk}}(Y_{j,nk} | Z_{j,nk}) \right) \right). \quad (2.9)$$

²For ease of notation we present probability mass function as $f_x(x)$.

From Eq. (2.7), Eq. (2.4) and Eq. (2.9) the posterior distribution is

$$f_{\mathbf{H}_{jj}, \mathbf{X}_j | \mathbf{Y}_j}(\mathbf{H}_{jj}, \mathbf{X}_j | \mathbf{Y}_j) = \frac{f_{\mathbf{Y}_j | \mathbf{Z}_j}(\mathbf{Y}_j | \mathbf{Z}_j) f_{\mathbf{H}_{jj}}(\mathbf{H}_{jj}) f_{\mathbf{X}_j}(\mathbf{X}_j)}{f_{\mathbf{Y}_j}(\mathbf{Y}_j)} \quad (2.10)$$

hence,

$$\begin{aligned} f_{\mathbf{H}_{jj}, \mathbf{X}_j | \mathbf{Y}_j}(\mathbf{H}_{jj}, \mathbf{X}_j | \mathbf{Y}_j) &\propto f_{\mathbf{Y}_j | \mathbf{Z}_j}(\mathbf{Y}_j | \mathbf{Z}_j) f_{\mathbf{H}_{jj}}(\mathbf{H}_{jj}) f_{\mathbf{X}_j}(\mathbf{X}_j) \\ &= \left(\prod_{n=1}^N \prod_{k=1}^K f_{Y_{j,nk} | Z_{j,nk}}(Y_{j,nk} | Z_{j,nk}) \right. \\ &\quad \left. \times \prod_{n=1}^N \prod_{k=1}^K f_{h_{jj,nk}}(h_{jj,nk}) \times \prod_{k=1}^K \prod_{t=1}^T f_{X_{j,kt}}(X_{j,kt}) \right). \end{aligned} \quad (2.11)$$

Given priors distributions, posterior distribution and observed data \mathbf{Y}_j , as defined in Eq. (2.7), Eq. (2.4) and Eq. (2.10), minimum mean-squared error (MMSE) estimates of \mathbf{H}_{jj} and \mathbf{X}_j can be computed by loopy belief propagation algorithm as we will analyze in Chap. 4.

Note that massive MIMO systems consist of a large number of received antennas, the complexity of joint detection and estimation algorithms is extremely notable, and needs to be evaluated.

Chapter 3

Closed-form Channel Estimation and Linear Data Detection

In this chapter closed-form estimators and detectors are provided in order to compute a minimum mean-squared error (MMSE) based estimate of \mathbf{H}_{jj} and \mathbf{X}_j^d in cell j of multi-cell massive MIMO system, given that pilot symbols \mathbf{X}_j^t are known. In the first algorithm presented in [7], BS j first uses \mathbf{X}_j^t to estimate \mathbf{H}_{jj} and then utilizes the estimated channel to evaluate data \mathbf{X}_j^d . In the second algorithm, a closed-form heuristic iterative procedure for joint estimation and data detection is proposed in order to minimize the estimation error.

3.1 Separate MMSE Estimation and Linear MMSE Detection

3.1.1 Channel Estimation

During uplink training phase, all UTs in each cell transmit the same set of orthogonal sequences. BSs use these training observations to compute the estimation matrix $\hat{\mathbf{H}}_{jj}$ of their local channels, i.e. channels between BS j and all UTs in cell j . For each UT k in cell j , BS correlates the observation in training phase

$$\mathbf{A}_j = [\mathbf{y}_{j,1} \ \mathbf{y}_{j,2} \ \dots \ \mathbf{y}_{j,T_t}] \in \mathbb{C}^{N \times T_t}, \text{ where } \mathbf{y}_{j,t} \in \mathbb{C}^N, \quad (3.1)$$

with the pilot sequence of UT k $\mathbf{x}_{j,k}^t \in \mathbb{C}^{1 \times T_t}$ divided by T_t in order to calculate $\mathbf{y}_{j,k}^{\text{tr}} \in \mathbb{C}^N$ which is the correlated received signal from UT k to BS j , i.e.,

$$\mathbf{y}_{j,k}^{\text{tr}} = \mathbf{A}_j \frac{(\mathbf{x}_{j,k}^t)^H}{T_t} = \mathbf{h}_{jj,k} + \sum_{l \neq j} \mathbf{h}_{jl,k} + \tilde{\mathbf{w}}_{j,k}, \quad (3.2)$$

where $\tilde{\mathbf{w}}_{j,k} \sim \mathcal{N}_{\mathbb{C}}(\tilde{\mathbf{w}}_{j,k}; \mathbf{0}, \frac{\sigma_w^2}{T_t} \mathbf{I}_N)$. Note that $\mathbf{x}_{j,k}^t (\mathbf{x}_{j,k}^t)^H = T_t$.

In [19] it is analyzed that in massive MIMO systems, measured wireless channels can be approached by theoretical channels with independent and identically distributed (i.i.d.) zero mean complex Gaussian coefficients, i.e., i.i.d. Rayleigh channels. Hence, the covariance matrix $\mathbf{R}_{jj,k}$ of $\mathbf{h}_{jj,k}$ is given by

$$\mathbf{R}_{jj,k} = \sigma_{h_{jj,k}}^2 \mathbf{I}_N. \quad (3.3)$$

Moreover, the covariance matrix $\mathbf{R}_{jl,k}$ of $\mathbf{h}_{jl,k}$ is given by

$$\mathbf{R}_{jl,k} = \sigma_{h_{jl,k}}^2 \mathbf{I}_N. \quad (3.4)$$

Therefore, based on $\mathbf{y}_{j,k}^{\text{tr}}$, the MMSE estimate $\hat{\mathbf{h}}_{jj,k}$ of $\mathbf{h}_{jj,k}$ is given as

$$\hat{\mathbf{h}}_{jj,k} = \mathbf{R}_{jj,k} \mathbf{Q}_{jj,k} \mathbf{y}_{j,k}^{\text{tr}} \quad (3.5)$$

where

$$\mathbf{Q}_{jj,k} = \left(\frac{\sigma_w^2}{T_t} \mathbf{I}_N + \sum_{l=1}^L \mathbf{R}_{jl,k} \right)^{-1}. \quad (3.6)$$

3.1.2 Linear Detection

Linear single-user detection is considered. In more detail, BS j compute the inner product between the observation vector $\mathbf{y}_{j,t}^d$ (observation in d-phase) and the linear filter $\mathbf{r}_{j,k}^{\text{MMSE}} \in \mathbb{C}^N$, with

$$\mathbf{r}_{j,k}^{\text{MMSE}} = \left(\hat{\mathbf{H}}_{jj} \hat{\mathbf{H}}_{jj}^H + \sigma_w^2 \mathbf{I}_N \right)^{-1} \hat{\mathbf{h}}_{jj,k} \quad (3.7)$$

for all T in order to estimate the data symbols $\mathbf{x}_{j,k}^d$ of UT k in its cell. Note that in MMSE linear filter proposed in [7] there exist the \mathbf{Z}_j^{ul} term, which is referred as the design parameter. The above term measures channel statistics of inter-cell UTs (inter-cell interference) and also the channel estimation errors (intra-cell channels). Practically, in multi-cell systems it is difficult for a BS to determine exactly the channel statistics of inter-cell UTs from remote cells. Hence, the above term is assumed to be zero. Moreover, for ease of notation the term which accounts for the intra-cell channel estimation errors is assumed to be zero. Therefore, the estimator in Eq. 3.7 may not minimize the mean squared errors; however, it is based on the MMSE equations and will be referred as MMSE thereafter.

3.1.3 Computation Complexity

Table 3.1: Matrices or Vectors Dimensions (SCED Algorithm)

Matrix / Vector	Dimensions
\mathbf{A}_j	$N \times T_t$
$(\mathbf{x}_{j,k}^t)^H$	$T_t \times 1$
$\mathbf{y}_{j,k}^{\text{tr}}$	$N \times 1$
$\mathbf{R}_{jj,k}$	$N \times N$
$\mathbf{Q}_{jj,k}$	$N \times N$
$\hat{\mathbf{h}}_{jj,k}$	$N \times 1$
$\hat{\mathbf{H}}_{jj}$	$N \times K$
$\mathbf{r}_{j,k}^{\text{MMSE}}$	$N \times 1$
\mathbf{Y}_j^d	$N \times T_d$

Computation complexity of separate MMSE channel estimation and linear MMSE data detection (SCED) algorithm based on Table. 3.1, is described by the following matrix multiplications

- Eq. 3.2: $N \times T_t \times K$ multiplications ($N \times T_t$ multiplications for K UTs)
- Eq. 3.6: $N^3 \times K$ multiplications (N^3 multiplications for K UTs)

- Eq. 3.5: $K \times N^2 \times (N + 2 \times T_t)$ multiplications¹
 - $N^3 \times K$ multiplications for $\mathbf{Q}_{jj,k}$ (N^3 multiplications for K UTs)
 - $N^2 \times T_t \times K$ multiplications for $\mathbf{Q}_{jj,k} \mathbf{y}_{j,k}^{\text{tr}}$ ($N^2 \times T_t$ multiplications for K UTs)
 - $N^2 \times T_t \times K$ multiplications for $\mathbf{R}_{jj,k}(\mathbf{Q}_{jj,k} \mathbf{y}_{j,k}^{\text{tr}})$ ($N^2 \times T_t$ multiplications for K UTs)
- Eq. 3.7: $N^2 \times (N + 2 \times K)$ multiplications
 - $N^2 \times K$ multiplications for $\hat{\mathbf{H}}_{jj} \hat{\mathbf{H}}_{jj}^H$
 - N^3 multiplications for $\left(\hat{\mathbf{H}}_{jj} \hat{\mathbf{H}}_{jj}^H + \sigma_w^2 \mathbf{I}_N \right)^{-1}$
 - $N^2 \times K$ multiplications for $\left(\hat{\mathbf{H}}_{jj} \hat{\mathbf{H}}_{jj}^H + \sigma_w^2 \mathbf{I}_N \right)^{-1} \hat{\mathbf{h}}_{jj,k}$ (N^2 multiplications for K UTs)
- For inner product between $\mathbf{y}_{j,t}^d$ and $\mathbf{r}_{j,k}^{\text{MMSE}}$: $N \times T_d \times K$ multiplications (N multiplications for K UTs and T_d symbols per UT).

In summary, complexity order is dominated by $K \times N^2 \times (N + 2 \times T_t)$ multiplications,² so $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t))$.

3.2 Iterative Estimation and Linear Detection

In this subsection, the heuristic iterative channel estimation and detection (JCED-Iterative) algorithm is presented. JCED-Iterative algorithm is inspired from separate MMSE channel estimation and linear MMSE data detection (SCED) algorithm presented in the above chapter. JCED-Iterative algorithm receives the detected symbols from SCED algorithm for all UTs

¹Note that $N \gg T_t$.

²From Eq. 3.5.

at d-phase duration as an input. Then, algorithm considers $\mathbf{X}_{\text{est},j} \in \mathbb{C}^{K \times T}$ as the new training phase, where

$$\mathbf{X}_{\text{est},j} = [\mathbf{X}_j^t \hat{\mathbf{X}}_j^d]. \quad (3.8)$$

Moore-Penrose inverse (pseudoinverse) matrix $\mathbf{X}_{\text{est},j}^+ \in \mathbb{C}^{T \times K}$, which described in [20], can be calculated as ³

$$\mathbf{X}_{\text{est},j}^+ = \mathbf{X}_{\text{est},j}^H \left(\mathbf{X}_{\text{est},j} \mathbf{X}_{\text{est},j}^H \right)^{-1} \quad (3.9)$$

and

$$\mathbf{X}_{\text{est},j}^+ = [\mathbf{x}_{\text{MP},t1} \ \mathbf{x}_{\text{MP},t2} \ \dots \ \mathbf{x}_{\text{MP},tK}] \in \mathbb{C}^{T \times K}, \text{ where } \mathbf{x}_{\text{MP},tk} \in \mathbb{C}^T. \quad (3.10)$$

Therefore,

$$\mathbf{X}_{\text{est},j} \mathbf{X}_{\text{est},j}^+ = \mathbf{I}_K. \quad (3.11)$$

3.2.1 Channel Estimation

For each UT k in BS j , JCED-Iterative algorithm follows the same procedure as SCED algorithm in order to estimate \mathbf{H}_{jj} and \mathbf{X}_j^d . In more detail,

$$\mathbf{y}_{\text{est},j,k}^{\text{tr}} = \mathbf{Y}_j \mathbf{x}_{\text{MP},j,k} = \mathbf{h}_{jj,k} + \mathbf{e}_{\text{inter}} + \bar{\mathbf{w}}_{j,k}, \quad (3.12)$$

where $\bar{\mathbf{w}}_{j,k} \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{w}}_{j,k}; \mathbf{0}, \|\mathbf{x}_{\text{MP},j,k}\|_2^2 \sigma_w^2 \mathbf{I}_N)^4$, $\mathbf{e}_{\text{inter}}$ accounts for the interference from all inter-cell UTs (not only the k UTs from other cells) and $\mathbf{y}_{\text{est},j,k}^{\text{tr}} \in \mathbb{C}^N$.

Hence, based on $\mathbf{y}_{\text{est},j,k}^{\text{tr}}$, the heuristic estimate $\hat{\mathbf{h}}_{jj,k}$ of $\mathbf{h}_{jj,k}$ is given as

$$\hat{\mathbf{h}}_{jj,k} = \mathbf{R}_{jj,k} \mathbf{Q}_{jj,k} \mathbf{y}_{\text{est},j,k}^{\text{tr}} \quad (3.13)$$

³Pseudoinverse matrix can be defined only when $K \leq T$.

⁴Proof: See Appendix A.

where $\mathbf{R}_{jj,k}$ remains as in Eq. 3.3 and

$$\mathbf{Q}_{jj,k} = \left(\|\mathbf{x}_{\text{MP},j,k}\|_2^2 \sigma_w^2 \mathbf{I}_N + \sum_{l=1}^L \sum_{k'=1}^K \mathbf{R}_{jl,k'} \right)^{-1}. \quad (3.14)$$

Note that Eq. 3.14 is a heuristic equation which also accounts for the interference from all inter-cell UTs (based on $\mathbf{e}_{\text{inter}}$ term in Eq. 3.12).

3.2.2 Linear Detection

Similarly to the SCED algorithm, BS j computes the inner product between the observation vector $\mathbf{y}_{j,t}^{\text{d}}$ and the linear filter $\mathbf{r}_{j,k} \in \mathbb{C}^N$ Eq. 3.7 for all T in order to estimate the symbols $\mathbf{x}_{j,k}^{\text{d}}$ of UT k in its cell.

3.2.3 Iterative Procedure

The new detected symbols for all UTs (as explained in Subsec. 3.2.2) are re-inputted to JCED-Iterative algorithm and then new values for channels and data symbols are calculated. This procedure is repeated until the algorithm converges or the maximum number of iterations is achieved. Fig. 3.1 shows a flow diagram of JCED-Iterative algorithm.

Convergence

JCED-Iterative algorithm converges when

$$\frac{\sum_{n=1}^N \sum_{t=T_t+1}^T |\hat{Z}_{j,nt}(\mu) - \hat{Z}_{j,nt}(\mu-1)|^2}{\sum_{n=1}^N \sum_{t=T_t+1}^T |\hat{Z}_{j,nt}(\mu)|^2} \leq \epsilon \quad (3.15)$$

where $\hat{\mathbf{Z}}_j = \hat{\mathbf{H}}_{jj} \hat{\mathbf{X}}_j$, μ is the current iteration and ϵ is a defined convergence parameter.

3.2.4 Computation Complexity

Computation complexity of heuristic Iterative based Channel Estimation and Detection (JCED-Iterative) algorithm based on Table. 3.2 is described by the

Table 3.2: Matrix Dimensions (SCED Algorithm)

Matrix	Dimensions
\mathbf{Y}_j	$N \times T$
$\mathbf{X}_{\text{est},j}^+$	$T \times K$
$\mathbf{x}_{\text{MP},j,k}$	$T \times 1$
$\mathbf{y}_{\text{est},j,k}^{\text{tr}}$	$N \times 1$
$\mathbf{R}_{jj,k}$	$N \times N$
$\mathbf{Q}_{jj,k}$	$N \times N$
$\hat{\mathbf{h}}_{jj,k}$	$N \times 1$
$\hat{\mathbf{H}}_{jj}$	$N \times K$
$\mathbf{r}_{j,k}$	$N \times 1$
\mathbf{Y}_j^{d}	$N \times T_{\text{d}}$

following matrix multiplications ⁵

- *Eq. 3.9*: $K^2 \times (K + 2 \times T)$ multiplications
 - $K^2 \times T$ multiplications for $\mathbf{X}_{\text{est},j} \mathbf{X}_{\text{est},j}^{\text{H}}$
 - K^3 multiplications for $\left(\mathbf{X}_{\text{est},j} \mathbf{X}_{\text{est},j}^{\text{H}}\right)^{-1}$
 - $T \times K^2$ multiplications for $\mathbf{X}_{\text{est},j}^+ = \mathbf{X}_{\text{est},j}^{\text{H}} \left(\mathbf{X}_{\text{est},j} \mathbf{X}_{\text{est},j}^{\text{H}}\right)^{-1}$.
- *Eq. 3.12*: $N \times T \times K$ multiplications ($N \times T$ multiplications for K UTs)
- *Eq. 3.14*: $N^3 \times K$ multiplications (N^3 multiplications for K UTs)
- *Eq. 3.13*: $K \times N^2 \times (N + 2 \times T)$ multiplications
 - $N^3 \times K$ multiplications for $\mathbf{Q}_{jj,k}$ (N^3 multiplications for K UTs)
 - $N^2 \times T \times K$ multiplications for $\mathbf{Q}_{jj,k} \mathbf{y}_{\text{est},j,k}^{\text{tr}}$ ($N^2 \times T_{\text{t}}$ multiplications for K UTs)

⁵For more details check Subsec. 3.1.3.

- $N^2 \times T \times K$ multiplications for $\mathbf{R}_{jj,k}(\mathbf{Q}_{jj,k}\mathbf{y}_{\text{est},j,k}^{\text{tr}})$ ($N^2 \times T$ multiplications for K UTs)
- *Eq. 3.7:* $N^2 \times (N + 2 \times K)$ multiplications
 - $N^2 \times K$ multiplications for $\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^H$
 - N^3 multiplications for $\left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^H + \sigma_w^2\mathbf{I}_N\right)^{-1}$
 - $N^2 \times K$ multiplications for $\left(\hat{\mathbf{H}}_{jj}\hat{\mathbf{H}}_{jj}^H + \sigma_w^2\mathbf{I}_N\right)^{-1}\hat{\mathbf{h}}_{jj,k}$ (N^2 multiplications for K UTs)
- *For inner product between $\mathbf{y}_{j,t}^d$ and $\mathbf{r}_{j,k}$:* $N \times T_d \times K$ multiplications (N multiplications for K UTs and T_d symbols per UT).

In summary, complexity order is dominated by $K \times N^2 \times (N + 2 \times T)$ multiplications⁶ per iteration, so $\mathcal{O}(\mu_{\text{con}} \times K \times N^2 \times (N + 2 \times T))$ where μ_{con} is the number of iteration that JCED-Iterative algorithm converges. However, as described above in this section, JCED-Iterative algorithm uses as initialization the detected symbols for all UTs at d-phase duration, that emerged from SCED algorithm. Computation complexity of these multiplications in SCED algorithm is $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t))$. Hence, the complexity order of JCED-Iterative algorithm is $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t) + \mu_{\text{con}} \times K \times N^2 \times (N + 2 \times T)) = \mathcal{O}\left(K \times N^2 \times ((\mu_{\text{con}} + 1) \times N + 2 \times (T_t + T))\right)$, where μ_{con} is the number of iteration that the JCED-Iterative algorithm converges.

⁶From Eq. 3.13.

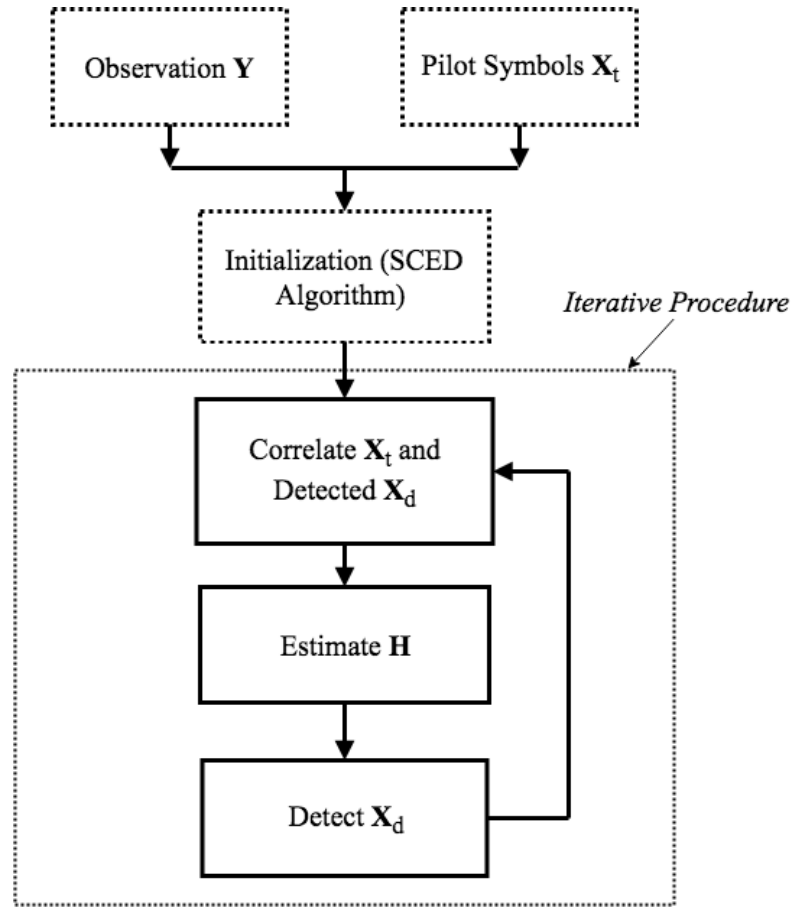


Figure 3.1: Flow diagram of JCED-Iterative algorithm.

Chapter 4

Bilinear Generalized Approximate Message Passing for MMSE Joint Data and Channel Estimation

In this chapter a graphical model approach is proposed, in order to compute an approximation of the minimum mean-squared error (MMSE) estimate of \mathbf{H}_{jj} , \mathbf{X}_j^d in cell j of multi-cell massive MIMO system. In more detail, we aim to compute means of the marginal posteriors $f_{\mathbf{H}_{jj}|\mathbf{Y}_j}(\mathbf{H}_{jj}|\mathbf{Y}_j)$ and $f_{\mathbf{X}_j|\mathbf{Y}_j}(\mathbf{X}_j|\mathbf{Y}_j)$ given the distributions of \mathbf{H}_{jj} , \mathbf{X}_j , $\mathbf{Y}_j|\mathbf{Z}_j$ and pilot symbols \mathbf{X}_j^t . In joint channel and data estimation scheme, BS j estimates both \mathbf{H}_{jj} and \mathbf{X}_j^d , in contrast to pilot-only scheme which the BS j first uses \mathbf{X}_j^t to estimate \mathbf{H}_{jj} and then uses the estimated channel to estimate data \mathbf{X}_j^d . For ease of notation, since we aim to estimate the data and the channel in cell j only, in this chapter we refer to \mathbf{H}_{jj} as \mathbf{H} , \mathbf{X}_j as \mathbf{X} , \mathbf{X}_j^t as \mathbf{X}_t and \mathbf{X}_j^d as \mathbf{X}_d .

In a graphical model, random variables are represented by nodes and relationships between these random variables represented by edges. Factor graph is a bipartite graph that contains two different classes of nodes, the factor nodes and the variable nodes. Factor graphs represent graphically the factorization of a global function into a product of local sub-functions. Using loopy belief propagation we can approximate the above means of the marginals [13]. More specifically, with sum-product algorithm (SPA) [14] the beliefs of the random variables, that propagate among the nodes of the factor graph until they converge, can be computed. The SPA algorithm operates by passing real valued functions, which are called messages, along the edges

between the factor and variable nodes. The belief that is transferred by a variable node to a factor node is computed as the product of all the incoming beliefs except the belief that variable node receives from the factor node that wants to transmit the message. On the other side, the belief that is transferred by a factor node to a variable node is computed as the integral of the product of the factor of all the incoming beliefs except the belief that factor node receives from the variable node that will transmit the outgoing message. The product of all the incoming beliefs in a variable node, gives the posterior distribution function (pdf) for that variable.

However, the computational complexity of the SPA in high-dimensional inference problems is infeasible. Generalized approximate message passing algorithm (GAMP) developed in [21], solves the generalized problem using a tractable method to approximate the marginal posterior of \mathbf{X} . In [13], authors derived an algorithm called bilinear approximate message passing algorithm, that uses GAMP-like approximations which are based on Taylor-series approximations and on central limit theorem arguments to compute means of the marginal posteriors of $f_{\mathbf{H}|\mathbf{Y}}(\mathbf{H}|\mathbf{Y})$ and $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{X}|\mathbf{Y})$.

4.1 Factor Graph

As we show in Sec. 2.2 at Eq. 2.11 posterior distribution $f_{\mathbf{H},\mathbf{X}|\mathbf{Y}}(\mathbf{H},\mathbf{X}|\mathbf{Y})$ can be written as a product of distributions $f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\mathbf{Z})$, $f_{\mathbf{H}}(\mathbf{H})$ and $f_{\mathbf{X}}(\mathbf{X})$. Therefore, the posterior distribution $f_{\mathbf{H},\mathbf{X}|\mathbf{Y}}(\mathbf{H},\mathbf{X}|\mathbf{Y})$ can be represented with a factor graph as it is shown in Fig. 4.1 where factor graphs for different parameters are represented. $f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\mathbf{Z})$, $f_{\mathbf{H}}(\mathbf{H})$ and $f_{\mathbf{X}}(\mathbf{X})$ are represented by the factor nodes which are designed with squares, while channel \mathbf{H} and transmitted signals \mathbf{X} are represented by the variable nodes, which are designed with circles. SPA algorithm, can be used to compute the marginal posterior distributions $f_{\mathbf{H}|\mathbf{Y}}(\mathbf{H}|\mathbf{Y})$ and $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{X}|\mathbf{Y})$.

As considered in Sec. 2.2, $\mathbf{Z} = \mathbf{H}\mathbf{X}$, with \mathbf{H} and \mathbf{X} unknown. In Sec. 4.2.2 an estimator for \mathbf{Z} is described. The same concept can be easily applied to estimate \mathbf{H} and \mathbf{X} . But, \mathbf{H} and \mathbf{X} are both unknown, so the complexity of estimator is still high. Problem can be divided into two separately problems,

the estimation of \mathbf{H} given \mathbf{X} and the estimation of \mathbf{H} given \mathbf{X} .

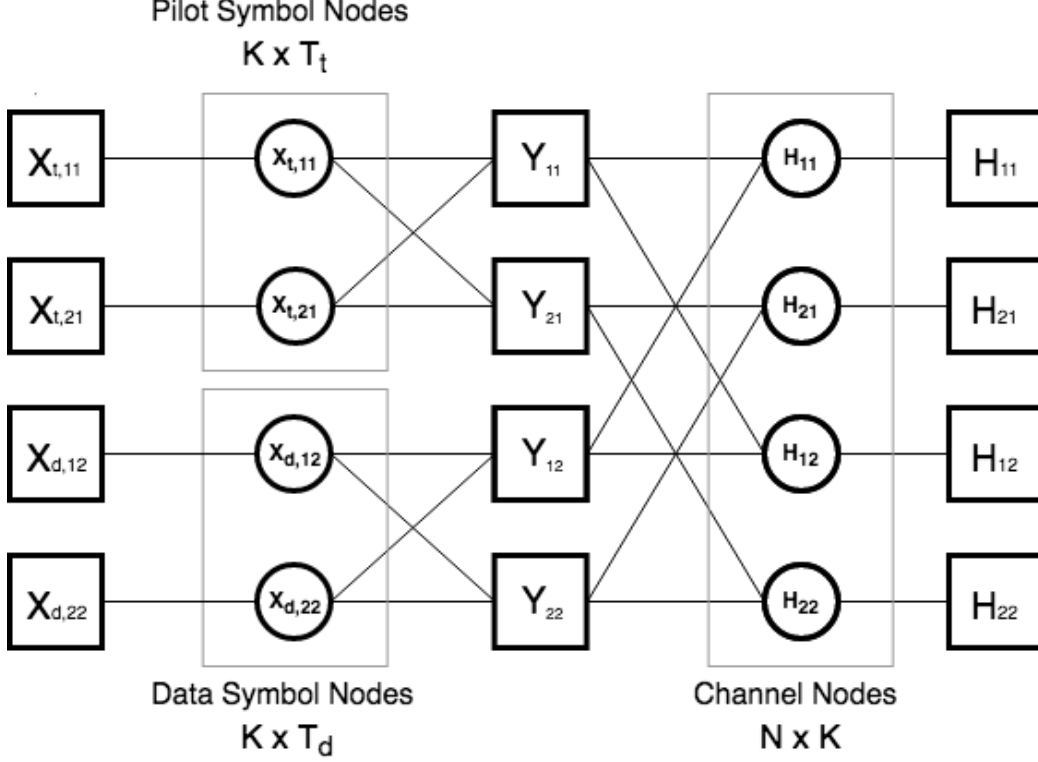


Figure 4.1: Graphical illustration of factor graph for MC-JCD-GAMP algorithm, where $N = 2$, $K = 2$, $T_t = 1$ and $T_d = 1$.

4.2 MC-JCD-GAMP Algorithm

An algorithm for joint channel and data estimation based on GAMP (JCD-GAMP) algorithm is proposed in [12]. We modify the above algorithm (to MC-JCD-GAMP) in order to use it in a multi-cell massive MIMO network, where interferences are caused by adjacent cells. Note that initialization phase, which is extremely important factor for the convergence of the algorithm, is changed compared to the algorithm that is proposed in [12]. Furthermore, changes were done at the estimator of \mathbf{Z} . MC-JCD-GAMP follows the same structure as Bi-GAMP [13] except the steps with pilot sym-

bols \mathbf{X}_t (orthogonal pilot symbols). In Algorithm 1 below, changes between JCD-GAMP and MC-JCD-GAMP are highlighted with blue color.

The steps in Algorithm 1 can be analyzed as follows. Lines 3-4, compute the estimate $\hat{\mathbf{P}}_t$ of the matrix product $\mathbf{Z}_t = \mathbf{H}\mathbf{X}_t$ and a corresponding set of elements-wise variances v_{nt}^p , $t \in \{1, \dots, T_t\}$. Lines 7-8, compute a “plug-in”¹ estimate $\bar{\mathbf{P}}_d$ of the matrix product $\mathbf{Z}_d = \mathbf{H}\mathbf{X}_d$ and a corresponding set of elements-wise variances \bar{v}_{nt}^p , $t \in \{T_t + 1, \dots, T\}$. Then in lines 9-10, “Onsager”² correction is applied to obtain the corresponding quantities $\hat{\mathbf{P}}_d$ and variances v_{nt}^p , $t \in \{T_t + 1, \dots, T\}$. Note that in lines 3-4 and 7-10 the algorithm follows the same procedure, but given that the pilot matrix \mathbf{X}_t is known, $v_{kt}^x = 0$, $t \in \{1, \dots, T_t\}$. Using the above quantities, in lines 12-13 the approximate marginal posterior means $\hat{\mathbf{Z}}$ and variances v_{nt}^z of \mathbf{Z} are computed. Lines 14-15 then use $\hat{\mathbf{Z}}$ and variances v_{nt}^z to compute the scaled residual $\hat{\mathbf{S}}$ and the inverse residual variances v_{nt}^s .³ Lines 16-17, then use the residual quantities to compute \hat{r}_{kt} and v_{kt}^r , where \hat{r}_{kt} can be defined as the observation $X_{d,kt}$ under an AWGN channel with zero mean and v_{kt}^r variance. Similarly, lines 18-19, compute the quantities \hat{q}_{nk} and v_{nk}^q , where \hat{q}_{nk} can be defined as the observation h_{nk} under an AWGN channel with zero mean and v_{nk}^q variance. Finally, lines 20-21 combine the terms \hat{r}_{kt} and v_{kt}^r with the prior $f_{\mathbf{X}_d}(\mathbf{X}_d)$ to compute the posterior mean $\hat{X}_{d,kt}$ and variance v_{kt}^x . Same procedure is followed in lines 22-23, for the calculation of the posterior mean \hat{H}_{nk} and variance v_{nk}^h .

For better understanding, Fig. 4.2 shows a representation of MC-JCD-GAMP algorithm. First, algorithm estimates at factor nodes means and variances of \mathbf{Z} from the observation \mathbf{Y} and incoming messages from variable nodes, if they exist. Then factor nodes send messages to variable nodes. After that, at variable/factor nodes \mathbf{X} , \mathbf{X} be observed under an AWGN channel, and the posterior means and variances of \mathbf{X} are computed. Parallel, at variable/factor nodes \mathbf{H} , \mathbf{H} be observed under an AWGN channel and the posterior means and variances of \mathbf{H} are computed. Both messages,

¹“Plug-in principle” is a simple method of estimating parameters from samples.

²For more details about “Onsager correction” in GAMP see [21].

³For more details about residual quantities in GAMP see [13].

Algorithm 1 MC-JCD-GAMP Algorithm

inputs: Observation \mathbf{Y} , pilot matrix \mathbf{X}_t , posterior distribution $f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{Y}|\mathbf{Z})$ and prior distributions $f_{\mathbf{H}}(\mathbf{H})$, $f_{\mathbf{X}}(\mathbf{X})$

outputs: Means of marginal posteriors $f_{\mathbf{H}|\mathbf{Y}}(\mathbf{H}|\mathbf{Y})$ and $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{X}|\mathbf{Y})$

initialization: $\mu = 1$, for $n = 1, \dots, N$, $k = 1, \dots, K$ and $t = 1, \dots, T$:

$\hat{s}_{nt}(0) = 0$, $\hat{H}_{nk}(1) = H_{\text{out},nk}$, $v_{nk}^h(1) \approx 0$ and for $t = T_t + 1, \dots, T$:

$\hat{X}_{kt}(1) = X_{\text{out},kt}$, $v_{kt}^x(1) = 1$

```

1: for  $\mu = 1, \dots, \mu_{\max}$  do
2:   for  $t = 1, \dots, T_t$  do
3:      $\forall n : v_{nt}^p(\mu) = \sum_{k=1}^K v_{nk}^h(\mu) |X_{kt}|^2$ 
4:      $\forall n : \hat{p}_{nt}(\mu) = \sum_{k=1}^K \hat{H}_{nk}(\mu) X_{kt} - \hat{s}_{nt}(\mu - 1) v_{nt}^p(\mu)$ 
5:   end for
6:   for  $t = T_t + 1, \dots, T$  do
7:      $\forall n : \bar{v}_{nt}^p(\mu) = \sum_{k=1}^K (|\hat{H}_{nk}(\mu)|^2 v_{kt}^x(\mu) + v_{nk}^h(\mu) |\hat{X}_{kt}(\mu)|^2)$ 
8:      $\forall n : \bar{p}_{nt}(\mu) = \sum_{k=1}^K \hat{H}_{nk}(\mu) \hat{X}_{kt}(\mu)$ 
9:      $\forall n : v_{nt}^p(\mu) = \bar{v}_{nt}^p(\mu) + \sum_{k=1}^K v_{nk}^h(\mu) v_{kt}^x(\mu)$ 
10:     $\forall n : \hat{p}_{nt}(\mu) = \bar{v}_{nt}^p(\mu) - \hat{s}_{nt}(\mu - 1) \bar{v}_{nt}^p(\mu)$ 
11:  end for

12:    $\forall n, t : v_{nt}^z(\mu) = \text{Var}\{Z_{nt} | \hat{p}_{nt}(\mu), v_{nt}^p(\mu)\}$ 
13:    $\forall n, t : \hat{Z}_{nt}(\mu) = \mathbb{E}\{Z_{nt} | \hat{p}_{nt}(\mu), v_{nt}^p(\mu)\}$ 

14:    $\forall n, t : v_{nt}^s(\mu) = (1 - v_{nt}^z(\mu) / v_{nt}^p(\mu)) / v_{nt}^p(\mu)$ 
15:    $\forall n, t : \hat{s}_{nt}(\mu) = (\hat{Z}_{nt}(\mu) - \hat{p}_{nt}(\mu)) / v_{nt}^p(\mu)$ 

16:    $\forall k, t : v_{kt}^r(\mu) = \left[ \sum_{n=1}^N |\hat{H}_{nk}(\mu)|^2 v_{nt}^s(\mu) \right]^{-1}$ 
17:    $\forall k, t : \hat{r}_{kt}(\mu) = \hat{X}_{kt}(\mu) (1 - v_{kt}^r(\mu) \sum_{n=1}^N v_{nk}^h(\mu) v_{nt}^s(\mu))$ 
    $+ v_{kt}^r(\mu) \sum_{n=1}^N \hat{H}_{nk}^*(\mu) \hat{s}_{nt}(\mu)$ 

18:    $\forall n, k : v_{nk}^q(\mu) = \left[ \sum_{t=1}^{T_t} |X_{kt}|^2 v_{nt}^s(\mu) + \sum_{t=T_t+1}^T |\hat{X}_{kt}(\mu)|^2 v_{nt}^s(\mu) \right]^{-1}$ 
19:    $\forall n, k : \hat{q}_{nk}(\mu) = \hat{H}_{nk}(\mu) (1 - v_{nk}^q(\mu) \sum_{t=1}^{T_t} v_{kt}^x(\mu) v_{nt}^s(\mu))$ 
    $+ v_{nk}^q(\mu) (\sum_{t=1}^{T_t} X_{kt}^* \hat{s}_{nt}(\mu) + \sum_{t=T_t+1}^T \hat{X}_{kt}^*(\mu) \hat{s}_{nt}(\mu))$ 

20:    $\forall k, t \in \{T_t + 1, \dots, T\} : v_{kt}^x(\mu + 1) = \text{Var}\{X_{kt} | \hat{r}_{kt}(\mu), v_{kt}^r(\mu)\}$ 
21:    $\forall k, t \in \{T_t + 1, \dots, T\} : \hat{X}_{kt}(\mu + 1) = \mathbb{E}\{X_{kt} | \hat{r}_{kt}(\mu), v_{kt}^r(\mu)\}$ 

22:    $\forall n, k : v_{nk}^h(\mu + 1) = \text{Var}\{H_{nk} | \hat{q}_{nk}(\mu), v_{nk}^q(\mu)\}$ 
23:    $\forall n, k : \hat{H}_{nk}(\mu + 1) = \mathbb{E}\{H_{nk} | \hat{q}_{nk}(\mu), v_{nk}^q(\mu)\}$ 
24:   if  $\frac{\sum_{n=1}^N \sum_{t=T_t+1}^T |\bar{p}_{nt}(\mu) - \bar{p}_{nt}(\mu-1)|^2}{\sum_{n=1}^N \sum_{t=T_t+1}^T |\bar{p}_{nt}(\mu)|^2} \leq \epsilon$  or  $\mu = \mu_{\max}$  then
25:     Stop
26:   end if
27: end for

```

from variable nodes \mathbf{X} and \mathbf{H} are sent to factor nodes. This procedure is repeated until the algorithm converges or the maximum number of iterations is achieved.

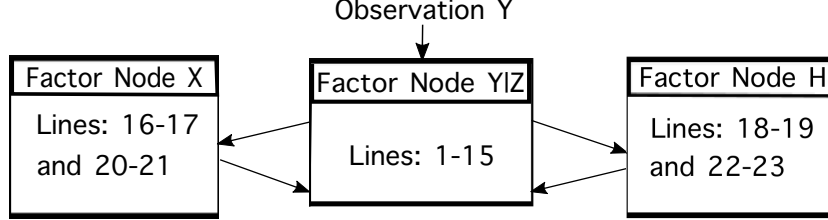


Figure 4.2: Representation of MC-JCD-GAMP algorithm.

4.2.1 Initialization

Initializations in MC-JCD-GAMP algorithm are extremely important for the convergence of the algorithm. As we describe in Ch. 3 during training phase, pilot contamination from adjacent cells is assumed, since the same set of orthogonal pilot sequences is reused in every cell. Given that the pilot symbols are known we can initialize the quantities $\hat{X}_{kt}(1)$ and $\hat{H}_{nk}(1)$ with the outputs $X_{\text{out},kt}$ and $H_{\text{out},nk}$ respectively, where $X_{\text{out},kt}$ and $H_{\text{out},nk}$ are computed from SCED algorithm that is presented in Sec. 3.1. Remaining terms \hat{s}_{nt} , v_{kt}^x and v_{nk}^h maintain their initial values as in [12].

4.2.2 Estimators

In this subsection, estimators that are used in lines 12-13, 20-21 and 22-23 are analyzed, based on [12]. As it is showed in MC-JCD-GAMP algorithm, in lines 12-13, 20-21 and 22-23 mean and variance of random variables Z_{nt} , X_{kt} and H_{nk} respectively, must be calculated. Hence, in this subsection closed-form estimators for the above variables are presented.

Estimator for \mathbf{Z}

Consider a SISO version of system Eq. 2.1 ($N = 1, T = 1$), $Y_j = Z_j + E_j + W_j$, where $W_j \sim \mathcal{N}_{\mathbb{C}}(W_j; 0, \sigma_w^2)$, $E_j \sim \mathcal{N}_{\mathbb{C}}(E_j; 0, \sum_{l \neq j} \sum_{k=1}^K \sigma_{h_{jl,k}}^2)$ ⁴ and $Z_j = H_{jj}X_j$. Hence, if we treat $U_j = E_j + W_j$ as additive Gaussian noise in Z_j , with $U_j \sim \mathcal{N}_{\mathbb{C}}(U_j; 0, \sigma_{u_j}^2)$ where $\sigma_{u_j}^2 = \sum_{l \neq j} \sum_{k=1}^K \sigma_{h_{jl,k}}^2 + \sigma_w^2$, the following SISO system can be considered

$$Y_j = Z_j + U_j. \quad (4.1)$$

For ease of notation, if j indicator removed

$$Y = Z + U \quad (4.2)$$

where,

$$f_{Y|Z}(Y|Z) = \frac{1}{\pi \sigma_u^2} e^{\frac{-|Y-Z|^2}{\sigma_u^2}}. \quad (4.3)$$

According to Bayes rule the posterior distribution of $Z|Y$ is given by

$$\begin{aligned} f_{H,X|Y}(H, X|Y) &= f_{Z|Y}(Z|Y) = \frac{f_{Y|Z}(Y|Z)f_Z(Z)}{f_Y(Y)} \\ &\propto f_{Y|Z}(Y|Z)f_Z(Z). \end{aligned} \quad (4.4)$$

Then, MMSE estimator for Z can be obtained by the posterior mean

$$\hat{Z} = \mathbb{E}[Z|Y] = \int z f(z|Y) dz. \quad (4.5)$$

Furthermore, to determine the above estimator assume that $Z \sim \mathcal{N}_{\mathbb{C}}(Z; \hat{p}, v^p)$. Hence, using the above distributions, can be calculated that

$$\begin{aligned} f_{Y|Z}(Y|Z)f_Z(Z) &= \mathcal{N}_{\mathbb{C}}(Z; Y, \sigma_u^2) \mathcal{N}_{\mathbb{C}}(Z; \hat{p}, v^p) \\ &= \mathcal{D} \cdot \mathcal{N}_{\mathbb{C}}\left(Z; \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 + v^p}, \frac{\sigma_u^2 v^p}{\sigma_u^2 + v^p}\right). \end{aligned} \quad (4.6)$$

⁴If QPSK modulation is assumed, ($\mathbb{E}[|X_j|^2] = 1$). For proof see Appendix B.

where $\mathcal{D} = \mathcal{N}_{\mathbb{C}}(0; Y - \hat{p}, \sigma_u^2 + v^p)$.⁵ Using Eq. 4.4 and Eq. 4.6 is computed that

$$f_{Z|Y}(Z|Y) = \mathcal{N}_{\mathbb{C}}\left(Z; \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 + v^p}, \frac{\sigma_u^2 v^p}{\sigma_u^2 + v^p}\right). \quad (4.7)$$

Finally, the estimator of Eq. 4.5, which is the mean of $f_{Z|Y}(Z|Y)$ is

$$\hat{Z} = \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 + v^p} = \hat{p} + \frac{v^p}{\sigma_u^2 + v^p} (Y - \hat{p}) \quad (4.8)$$

and the MSE of the above estimator, which is the variance of $f_{Y|Z}(Y|Z)$ is

$$v^z = \frac{\sigma_u^2 v^p}{\sigma_u^2 + v^p} = v^p - \frac{(v^p)^2}{\sigma_u^2 + v^p}. \quad (4.9)$$

Estimator for H

As defined in Subsec. 2.1.1, $f_{h_{n,k}}(h_{n,k}) \equiv \mathcal{N}_{\mathbb{C}}(h_{n,k}; 0, \sigma_{h_k}^2)$. Furthermore, it is analyzed that \hat{q}_{nk} term can be defined as the observation of h_{nk} under an AWGN channel with variance v_{nk}^q . For ease of notation, terms n and k are omitted in next equations. Therefore, the posterior distribution of h given observation \hat{q}_{nk} according to Bayes rule is

$$\begin{aligned} f_{h|\hat{q}}(h|\hat{q}) &= \frac{f_{\hat{q}|h}(\hat{q}|h)f_h(h)}{f_{\hat{q}}(\hat{q})} \\ &\propto f_{\hat{q}|h}(\hat{q}|h)f_h(h). \end{aligned} \quad (4.10)$$

Thereby, the MMSE estimator for h can be obtained by the posterior mean

$$\hat{h} = \mathbb{E}[h|\hat{q}] = \int h f_{h|\hat{q}}(h|\hat{q}) dh. \quad (4.11)$$

Then following the same procedure as in estimator for Z ,

$$f_{\hat{q}|h}(\hat{q}|h)f_h(h) = \mathcal{N}_{\mathbb{C}}(h; \hat{q}, v^q) \mathcal{N}_{\mathbb{C}}(h; 0, \sigma_h^2). \quad (4.12)$$

⁵Proof: See Appendix C.

Therefore, the mean of $f_{h|\hat{q}}(h|\hat{q})$, which is the MMSE estimator of h , is

$$\hat{h} = \frac{\sigma_h^2}{\sigma_h^2 + v^q} \hat{q} \quad (4.13)$$

and the MSE of the above estimator, which is the variance of $f_{h|\hat{q}}(h|\hat{q})$ becomes

$$v^h = \frac{\sigma_h^2 v^q}{\sigma_h^2 + v^q} = v^q - \frac{(v^q)^2}{\sigma_h^2 + v^q}. \quad (4.14)$$

Estimator for X_d

It is analyzed that \hat{r}_{kt} term can be defined as the observation of $X_{d,kt}$ under an AWGN channel with variance v_{kt}^r . For ease of notation, terms k and t are omitted in next equations. Hence, the posterior distribution of X_d given observation \hat{r}_{kt} according to Bayes rule is

$$\mathcal{P}(X_d) = \frac{f_{\hat{r}|X_d}(\hat{r}|X_d)\Pr(X_d)}{\sum_{X'_d} f_{\hat{r}|X'_d}(\hat{r}|X'_d)\Pr(X'_d)}. \quad (4.15)$$

where $f_{\hat{r}|X_d}(\hat{r}|X_d) = \mathcal{N}_{\mathbb{C}}(\hat{r}; X_d, v^r) = \mathcal{N}_{\mathbb{C}}(X_d; \hat{r}, v^r)$. Thereby, the estimator for X_d can be obtained by the posterior mean

$$\hat{X}_d = \mathbb{E}[X_d|\hat{r}] = \sum_{X_d \in \mathcal{B}} X_d \mathcal{P}(X_d) \quad (4.16)$$

QPSK modulation is considered, where 2×2 points are

$$\mathcal{B} = \frac{1}{\sqrt{2}} \times \left\{ 1 + j, -1 + j, 1 - j, -1 - j \right\} \quad (4.17)$$

and, $\Pr(X_d) = 1/4$ for $X_d \in \mathcal{B}$. Then, lines 22-23 can be calculated as ⁶

$$\hat{X}_d = \frac{1}{\sqrt{2}} \tanh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}v^r}\right) + j \frac{1}{\sqrt{2}} \tanh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}v^r}\right), \quad (4.18)$$

$$v^x = \frac{1}{2} \tanh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}v^r}\right) + \frac{1}{2} \tanh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}v^r}\right) - |\hat{X}_d|^2. \quad (4.19)$$

⁶Proof: See Appendix D.

4.2.3 Damping Factor

In the algorithm that is presented above in this section, for ease of interpretation, did not included the important damping modification. In this subsection, the damping factor $\beta \in (0, 1]$ that used to slow the evolution of certain variables is analyzed. Note that damping factor is extremely important for the convergence of the MC-JCD-GAMP algorithm.

In more detail, lines 2, 7, 8, 14 and 15 should be replaced with

$$v_{nt}^p(\mu) = \beta \left(\sum_{k=1}^K v_{nk}^h(\mu) |X_{kt}|^2 \right) + (1 - \beta) v_{nt}^p(\mu - 1) \quad (4.20)$$

$$\begin{aligned} \bar{v}_{nt}^p(\mu) &= \beta \left(\sum_{k=1}^K (|\hat{H}_{nk}(\mu)|^2 v_{kt}^x(\mu) + v_{nk}^h(\mu) |\hat{X}_{kt}(\mu)|^2) \right) \\ &\quad + (1 - \beta) \bar{v}_{nt}^p(\mu - 1) \end{aligned} \quad (4.21)$$

$$v_{nt}^p(\mu) = \beta \left(\bar{v}_{nt}^p(\mu) + \sum_{k=1}^K v_{nk}^h(\mu) v_{kt}^x(\mu) \right) + (1 - \beta) v_{nt}^p(\mu - 1) \quad (4.22)$$

$$v_{nt}^s(\mu) = \beta \left((1 - v_{nt}^z(\mu) / v_{nt}^p(\mu)) / v_{nt}^p(\mu) \right) + (1 - \beta) v_{nt}^s(\mu - 1) \quad (4.23)$$

$$\hat{s}_{nt}(\mu) = \beta \left((\hat{Z}_{nt}(\mu) - \hat{p}_{nt}(\mu)) / v_{nt}^p(\mu) \right) + (1 - \beta) \hat{s}_{nt}(\mu - 1) \quad (4.24)$$

respectively, and between lines 15-16 are inserted the following two lines

$$\bar{X}_{kt}(\mu) = \beta \hat{X}_{kt}(\mu) + (1 - \beta) \bar{X}_{kt}(\mu - 1) \quad (4.25)$$

$$\bar{H}_{nk}(\mu) = \beta \hat{H}_{nk}(\mu) + (1 - \beta) \bar{H}_{nk}(\mu - 1) \quad (4.26)$$

then, the $\bar{X}_{kt}(\mu)$ and $\bar{H}_{nk}(\mu)$ are used in places of $\hat{X}_{kt}(\mu)$ and $\hat{H}_{nk}(\mu)$ in lines 16-17 and 18-19 respectively.

4.2.4 Computation Complexity

Computation complexity of MC-JCD-GAMP algorithm is described by the following matrix multiplications per line per iteration

- *Lines 3-4:* $N \times K \times T_t$ multiplications
- *Line 7:* $2 \times N \times K \times T_d$ multiplications
- *Lines 8-9:* $N \times K \times T_d$ multiplications
- *Lines 10, 12-15:* $N \times T$ multiplications
- *Line 16:* $N \times K \times T$ multiplications
- *Line 17:* $2 \times N \times K \times T$ multiplications
- *Line 18:* $N \times K \times T_t + N \times K \times T_d$ multiplications
- *Line 19:* $N \times K \times T_d + N \times K \times T_t + N \times K \times T_d$ multiplications
- *Lines 20-21:* $N \times T_d$ multiplications
- *Lines 22-23:* $N \times K$ multiplications.

In summary, complexity order is dominated by $N \times K \times T$ multiplications per iteration, so $\mathcal{O}(\mu_{\text{con}} \times N \times K \times T)$ where μ_{con} is the number of iteration that MC-JCD-GAMP algorithm converges.

However, as described in Sec. 4.2.1, in initialization phase, SCED algorithm is used to initialize the terms of \hat{X}_{nt} and \hat{H}_{nk} . Computation complexity of these multiplications in SCED algorithm is $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t))$. Hence, the complexity order of MC-JCD-GAMP algorithm is $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t) + \mu_{\text{con}} \times (N \times K \times T)) = \mathcal{O}(K \times N^2 \times (N + 2 \times T_t))$.

4.3 MC-JCD-GAMP-D Algorithm

In this section, MC-JCD-GAMP algorithm is modified to MC-JCD-GAMP-D algorithm, i.e. in the time duration $T_d = T - T_t$. In more detail, pilot

symbols X_t are used only in initialization phase, similarly to MC-JCD-GAMP algorithm Subsec. 4.2.1. Then, the approximate message passing algorithm is applied in d-phase only, hence observation \mathbf{Y}_d is used, i.e.

$$\mathbf{Y}_d = [\mathbf{y}_{T_t+1} \ \mathbf{y}_{T_t+2} \ \dots \ \mathbf{y}_T] \in \mathbb{C}^{N \times T_d}, \text{ where } \mathbf{y}_t \in \mathbb{C}^N. \quad (4.27)$$

Fig. 4.3 shows a flow diagram of the above algorithm.

As we show in Sec. 2.2 at Eq. 2.11 posterior distribution $f_{\mathbf{H}, \mathbf{X} | \mathbf{Y}}(\mathbf{H}, \mathbf{X} | \mathbf{Y})$ can be written as a product of distributions $f_{\mathbf{Y} | \mathbf{Z}}(\mathbf{Y} | \mathbf{Z})$, $f_{\mathbf{H}}(\mathbf{H})$ and $f_{\mathbf{X}}(\mathbf{X})$. In this section, terms \mathbf{Y} , \mathbf{Z} , \mathbf{X} and $f_{\mathbf{X}}(\mathbf{X})$ are replaced by \mathbf{Y}_d , \mathbf{Z}_d , \mathbf{X}_d and $f_{\mathbf{X}_d}(\mathbf{X}_d)$, respectively. Therefore, the posterior distribution $f_{\mathbf{H}, \mathbf{X}_d | \mathbf{Y}_d}(\mathbf{H}, \mathbf{X}_d | \mathbf{Y}_d)$ is

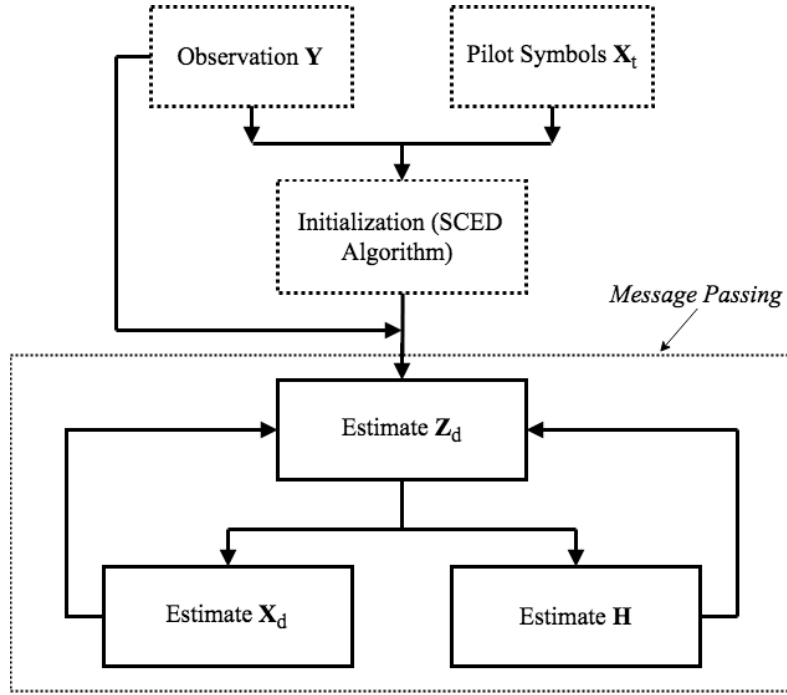


Figure 4.3: Flow diagram of MC-JCD-GAMP-D algorithm.

$$\begin{aligned}
f_{\mathbf{H}, \mathbf{X}_d | \mathbf{Y}_d}(\mathbf{H}, \mathbf{X}_d | \mathbf{Y}_d) &\propto f_{\mathbf{Y}_d | \mathbf{Z}_d}(\mathbf{Y}_d | \mathbf{Z}_d) f_{\mathbf{H}}(\mathbf{H}) f_{\mathbf{X}_d}(\mathbf{X}_d) \\
&= \prod_{n=1}^N \left(\prod_{k=1}^K \left(f_{Y_{d,nk} | Z_{d,nk}}(Y_{d,nk} | Z_{d,nk}) \right) \right) \\
&\quad \times \prod_{n=1}^N \left(\prod_{k=1}^K \left(f_{h_{nk}}(h_{nk}) \right) \right) \\
&\quad \times \prod_{k=1}^K \left(\prod_{t=T_t+1}^T \left(f_{X_{kt}}(X_{kt}) \right) \right). \tag{4.28}
\end{aligned}$$

The factor graph that represents the Eq. 4.28 is shown in Fig. 4.4. As showed, the variable nodes for \mathbf{X}_t do not exist anymore. $f_{\mathbf{Y}_d | \mathbf{Z}_d}(\mathbf{Y}_d | \mathbf{Z}_d)$, $f_{\mathbf{H}}(\mathbf{H})$ and $f_{\mathbf{X}_d}(\mathbf{X}_d)$ is represented by the factor nodes which are designed with squares, while \mathbf{H} and \mathbf{X}_d are represented by the variable nodes, which are designed with cycles. SPA algorithm, can be used to compute the marginal posterior distributions $f_{\mathbf{H} | \mathbf{Y}_d}(\mathbf{H} | \mathbf{Y}_d)$ and $f_{\mathbf{X}_d | \mathbf{Y}_d}(\mathbf{X}_d | \mathbf{Y}_d)$. Algorithm 2 (Alg. 2) describes the steps of the MC-JCD-GAMP-D algorithm. Steps are exactly the same as in the Bi-GAMP algorithm in [13]. Initialization, damping steps and estimation functions (lines 6-7, 14-15, 16-17) are the same as in the algorithm presented in Sec. 4.2.

4.3.1 Computation Complexity

Computation complexity of the MC-JCD-GAMP-D algorithm is described by the following matrix multiplications per line per iteration

- *Line 2*: $2 \times N \times K \times T_d$ multiplications
- *Lines 3-4*: $N \times K \times T_d$ multiplications
- *Lines 5-9*: $N \times T_d$ multiplications
- *Line 10*: $N \times K \times T_d$ multiplications
- *Line 11*: $2 \times N \times K \times T_d$ multiplications

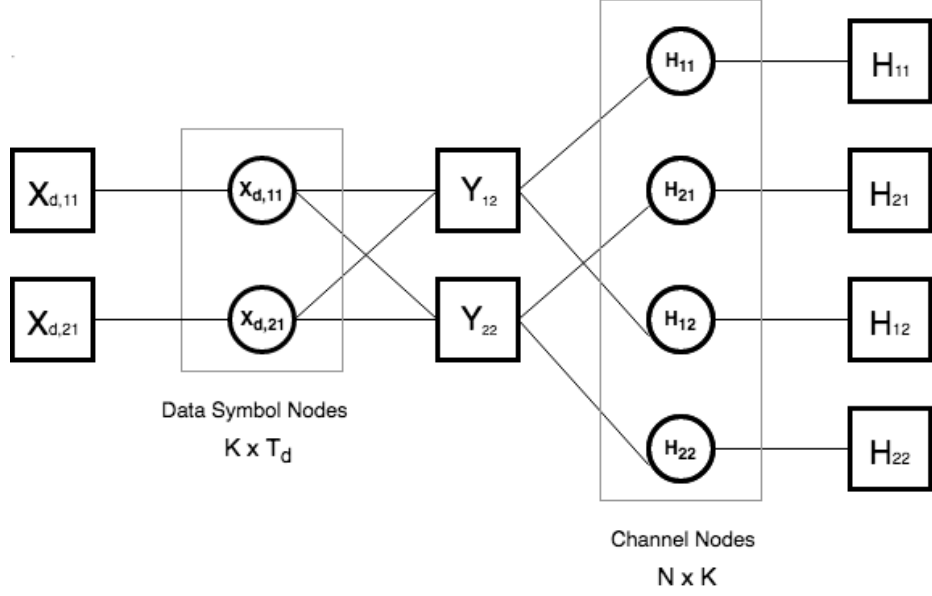


Figure 4.4: Graphical illustrator of factor graph for the MC-JCD-GAMP-D algorithm, where $N = 2$, $K = 2$ and $T_d = 1$.

- *Line 12:* $N \times K \times T_d$ multiplications
- *Line 13:* $2 \times N \times K \times T_d$ multiplications
- *Lines 20-21:* $N \times T_d$ multiplications
- *Lines 22-23:* $N \times K$ multiplications.

In conclusion, complexity order is dominated by $N \times K \times T_d$ multiplications per iteration, therefore $\mathcal{O}(\mu_{\text{con}} \times N \times K \times T_d)$ where μ_{con} is the number of iteration that MC-JCD-GAMP-D algorithm converges.

However, as described in Subsec. 4.2.4, in initialization phase, the computation complexity is $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t))$. Therefore, the complexity order of MC-JCD-GAMP-D algorithm is $\mathcal{O}(K \times N^2 \times (N + 2 \times T_t) + \mu_{\text{con}} \times N \times K \times T_d) = \mathcal{O}(K \times N^2 \times (N + 2 \times T_t))$.

Algorithm 2 MC-JCD-GAMP-D Algorithm

inputs: Observation \mathbf{Y}_d , pilot matrix \mathbf{X}_t (in initialization phase), posterior distribution $f_{\mathbf{Y}_d|\mathbf{Z}_d}(\mathbf{Y}_d|\mathbf{Z}_d)$ and prior distributions $f_{\mathbf{H}}(\mathbf{H})$, $f_{\mathbf{X}_d}(\mathbf{X}_d)$

outputs: Means of marginal posteriors $f_{\mathbf{H}|\mathbf{Y}_d}(\mathbf{H}|\mathbf{Y}_d)$ and $f_{\mathbf{X}_d|\mathbf{Y}_d}(\mathbf{X}_d|\mathbf{Y}_d)$

initialization: $\mu = 1$, for $n = 1, \dots, N$, $k = 1, \dots, K$ and $t = T_t + 1, \dots, T$:
 $\hat{s}_{nt}(0) = 0$, $\hat{X}_{kt}(1) = X_{\text{out},kt}$, $v_{kt}^x(1) = 1$, $\hat{H}_{nk}(1) = H_{\text{out},nk}$, $v_{nk}^h(1) \approx 0$

- 1: **for** $\mu = 1, \dots, \mu_{\max}$ **do**
- 2: $\forall n : \bar{v}_{nt}^p(\mu) = \sum_{k=1}^K (|\hat{H}_{nk}(\mu)|^2 v_{kt}^x(\mu) + v_{nk}^h(\mu) |\hat{X}_{kt}(\mu)|^2)$
- 3: $\forall n : \bar{p}_{nt}(\mu) = \sum_{k=1}^K \hat{H}_{nk}(\mu) \hat{X}_{kt}(\mu)$
- 4: $\forall n : v_{nt}^p(\mu) = \bar{v}_{nt}^p(\mu) + \sum_{k=1}^K v_{nk}^h(\mu) v_{kt}^x(\mu)$
- 5: $\forall n : \hat{p}_{nt}(\mu) = \bar{v}_{nt}^p(\mu) - \hat{s}_{nt}(\mu - 1) \bar{v}_{nt}^p(\mu)$
- 6: $\forall n, t : v_{nt}^z(\mu) = \text{Var}\{Z_{nt}|\hat{p}_{nt}(\mu), v_{nt}^p(\mu)\}$
- 7: $\forall n, t : \hat{Z}_{nt}(\mu) = \mathbb{E}\{Z_{nt}|\hat{p}_{nt}(\mu), v_{nt}^p(\mu)\}$
- 8: $\forall n, t : v_{nt}^s(\mu) = (1 - v_{nt}^z(\mu)/v_{nt}^p(\mu))/v_{nt}^p(\mu)$
- 9: $\forall n, t : \hat{s}_{nt}(\mu) = (\hat{Z}_{nt}(\mu) - \hat{p}_{nt}(\mu))/v_{nt}^p(\mu)$
- 10: $\forall k, t : v_{kt}^r(\mu) = \left[\sum_{n=1}^N |\hat{H}_{nk}(\mu)|^2 v_{nt}^s(\mu) \right]^{-1}$
- 11: $\forall k, t : \hat{r}_{kt}(\mu) = \hat{X}_{kt}(\mu) (1 - v_{kt}^r(\mu) \sum_{n=1}^N v_{nk}^h(\mu) v_{nt}^s(\mu))$
 $\quad + v_{kt}^r(\mu) \sum_{n=1}^N \hat{H}_{nk}^*(\mu) \hat{s}_{nt}(\mu)$
- 12: $\forall n, k : v_{nk}^q(\mu) = \left[\sum_{t=T_t+1}^T |\hat{X}_{kt}(\mu)|^2 v_{nt}^s(\mu) \right]^{-1}$
- 13: $\forall n, k : \hat{q}_{nk}(\mu) = \hat{H}_{nk}(\mu) (1 - v_{nk}^q(\mu) \sum_{t=1+T_t}^T v_{kt}^x(\mu) v_{nt}^s(\mu))$
 $\quad + v_{nk}^q(\mu) (\sum_{t=T_t+1}^T \hat{X}_{kt}^*(\mu) \hat{s}_{nt}(\mu))$
- 14: $\forall k, t : v_{kt}^x(\mu + 1) = \text{Var}\{X_{kt}|\hat{r}_{kt}(\mu), v_{kt}^r(\mu)\}$
- 15: $\forall k, t : \hat{X}_{kt}(\mu + 1) = \mathbb{E}\{X_{kt}|\hat{r}_{kt}(\mu), v_{kt}^r(\mu)\}$
- 16: $\forall n, k : v_{nk}^h(\mu + 1) = \text{Var}\{H_{nk}|\hat{q}_{nk}(\mu), v_{nk}^q(\mu)\}$
- 17: $\forall n, k : \hat{H}_{nk}(\mu + 1) = \mathbb{E}\{H_{nk}|\hat{q}_{nk}(\mu), v_{nk}^q(\mu)\}$
- 18: **if** $\frac{\sum_{n=1}^N \sum_{t=T_t+1}^T |\bar{p}_{nt}(\mu) - \bar{p}_{nt}(\mu-1)|^2}{\sum_{n=1}^N \sum_{t=T_t+1}^T |\bar{p}_{nt}(\mu)|^2} \leq \epsilon$ **or** $\mu = \mu_{\max}$ **then**
- 19: **Stop**
- 20: **end if**
- 21: **end for**

Chapter 5

Simulation Results

In this chapter we compare through simulations the performance of algorithms (SCED, SCED-Iterative, MC-JCD-GAMP and MC-JCD-GAMP-D) presented in Ch. 3 and Ch. 4. A hexagonal cellular network is considered based on [15] with $L = 7$ cells and frequency reuse factor = 1. Each cell contains uniformly distributed UTs around BS (which is placed in the center of each cell). Interference from adjacent cells is considered. Figures in this chapter depict various measurement results, such as the average Symbol Error Rate (SER), concerning only the intra-cell UTs located in the central BS of a topology, while the inter-cell UTs are considered only as interferers. Moreover, the simulation results are calculated by averaging Monte-Carlo runs with different channel realizations.

Realistic assumptions for multi-cell massive MIMO are considered based on Long Term Evolution (LTE). Cell radius is assumed to be 1000m, path-loss exponent $\alpha = 3.7$, antenna gain at BS $G_{\text{BS}} = 16$ dBi and antenna gain at each UT $G_{\text{UT}} = 2.15$ dBi. Furthermore, bandwidth of 10 MHz is considered, while the uplink carrier frequency is assumed to be at $f_{\text{carrier}} = 1950$ MHz. Free space wavelength λ is $\frac{v}{f_{\text{carrier}}} = \frac{3 \times 10^8}{f_{\text{carrier}}}$ m where v is the speed of light, while the reference distance $d_0 = 100$ m. In addition, Noise Figure (NF) is assumed to be 4 dB. QPSK modulation is considered for transmit data, and the convergence parameter $\epsilon = 10^{-6}$. Fig. 5.1 shows a random topology with $K = 10$ UTs in each cell. Note that received SNR at BS from intra-cell UT k is defined as

$$\text{SNR}_k = \frac{\sigma_{h_k}^2}{\sigma_w^2} \quad (5.1)$$

where $\sigma_{h_k}^2$ is the received signal power at BS from UT k and σ_w^2 the power of noise at BS. Furthermore, normalized root mean square error (NRMSE)

is considered as the metric for channel estimation of intra-cell links where

$$\text{NRMSE} = \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_F}{\|\mathbf{H}\|_F}. \quad (5.2)$$

At first and second scenario, SER and NRMSE refer to the average SER and NRMSE across all intra-cell UTs, respectively. At third scenario, SER and NRMSE refer to SER and NRMSE of intra-cell UT k , respectively.

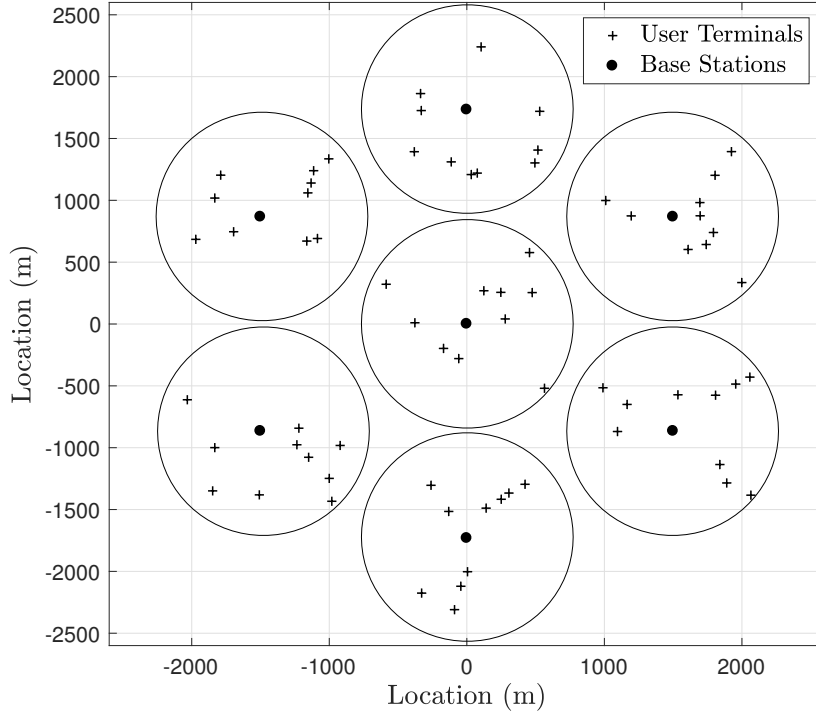


Figure 5.1: Topology of 7-cell hexagonal system with $K = 10$ UTs uniformly distributed around each BS.

5.1 Impact of Damping Factor

In this section the effect of damping factor in MC-JCD-GAMP, MC-JCD-GAMP-D algorithms is shown under different transmit power scenarios. The simulations are conducted for $N = 64$ antennas at central BS, $K = 10$ UTs

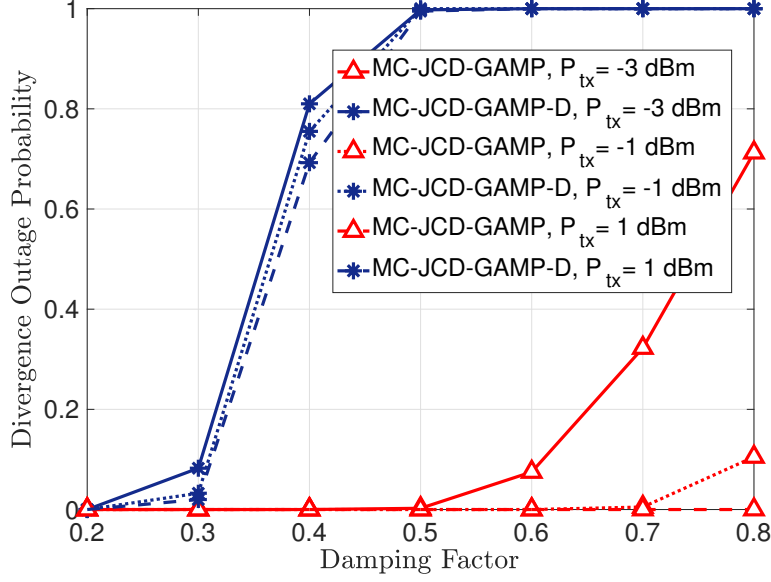


Figure 5.2: Divergence outage probability versus damping factor, for $N = 64$, $K = 10$ and $T_d = 240$.

per cell and $T = 250$ symbols ($T_t = 10$ pilot symbols, $T_d = 240$ data symbols). Moreover, the transmit power (P_{tx} or P^{transmit}) of intra-cell UTs is assumed to be -3 dBm, -1 dBm and 1 dBm, while the transmit power of inter-cell UTs is set to 2 dBm. In Fig. 5.2, Fig. 5.3 it is shown that MC-JCD-GAMP-D algorithm needs smaller damping factor than MC-JCD-GAMP algorithm in order to converge to an accepted estimation. Moreover, it is observed that the damping factor in the MC-JCD-GAMP-D algorithm is more critical a parameter than in MC-JCD-GAMP algorithm. The above observation can be explained by the fact that multiplications in MC-JCD-GAMP algorithm use pilot symbols X_t (which are constant and known symbols) in message passing at all iterations until converge. Hence, more stability is provided than the MC-JCD-GAMP-D algorithm, in which X_t symbols do not exist in the multiplications. Furthermore, as damping factor decreases, MC-JCD-GAMP algorithm converges to an accepted estimation with higher certainty. However, in most cases more iterations are needed to achieve an accepted estimation with such high certainty in comparison with having higher damping factor. More specifically, if the damping factor increases, MC-JCD-GAMP

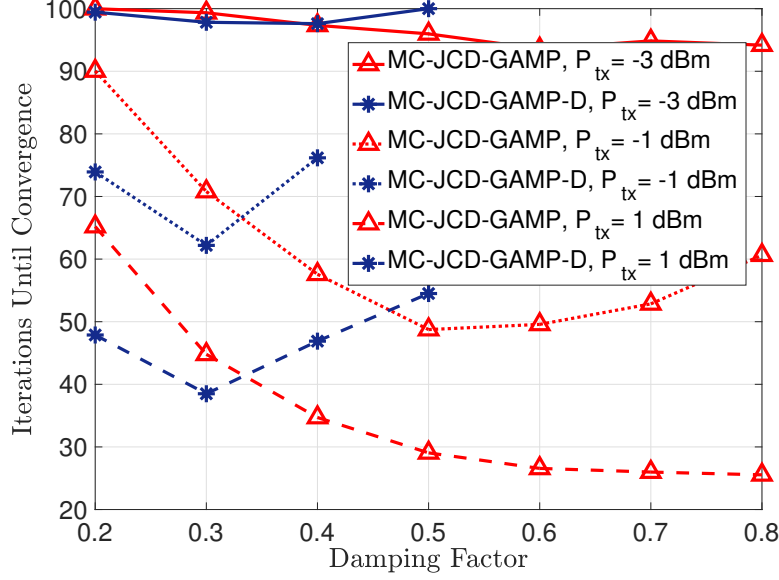


Figure 5.3: Iterations until convergence versus damping factor, for $N = 64$, $K = 10$ and $T_d = 240$.

algorithm may converge in fewer iterations (not in all cases) but also may diverge with high probability. Moreover, MC-JCD-GAMP-D algorithm seems to converge with damping factor smaller than 0.5, with best performance when damping factor is equal to 0.2 or 0.3. In general case, damping factor must be smaller than 0.7 so that BiGAMP-based algorithms can converge with high probability. Therefore, damping factor is an important parameter and needs further research in order to select the optimized value in each scenario. In the next sections, damping factor is selected between 0.2 and 0.5 for the MC-JCD-GAMP-D algorithm, while for the MC-JCD-GAMP algorithm damping factor is selected between 0.5 and 0.7.

5.2 First Scenario (64 or 128 Received Antennas, 10 UTs per Cell)

The first set of simulations is conducted for $N = 64$ antennas at central BS, $K = 10$ UTs per cell and $T = 250$ symbols ($T_t = 10$ pilot symbols, $T_d = 240$

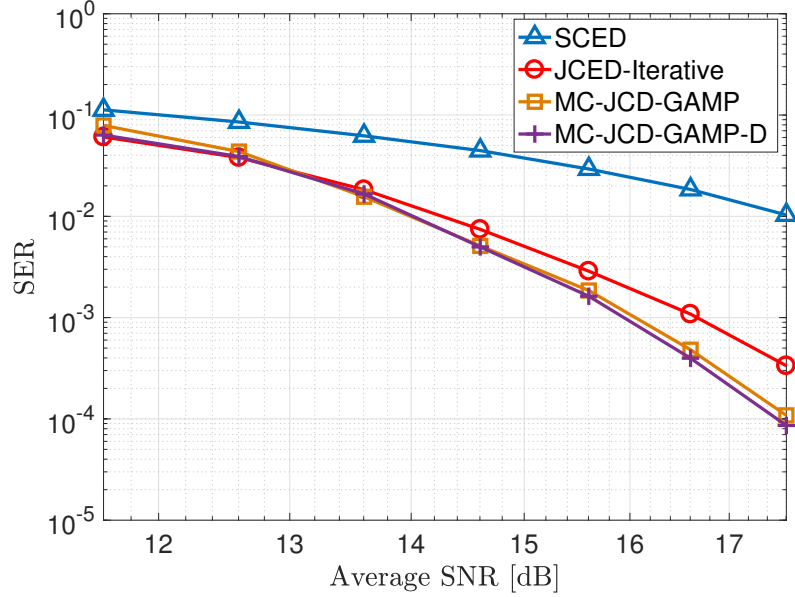


Figure 5.4: Symbol Error Rate versus average received SNR, for $N = 64$, $K = 10$ and $T_d = 240$.

data symbols). In order to obtain different values for SNR and SINR, the transmit power of intra-cell UTs is varied from -4 dBm to 2 dBm, while the transmit power of inter-cell UTs is held constant at 2 dBm.

In Fig. 5.4, Fig. 5.5, Fig. 5.6 and Fig. 5.7 SER performances of SCED, JCED-Iterative, MC-JCD-GAMP, MC-JCD-GAMP-D algorithms are shown versus average, minimum and maximum received SNR at BS from intra-cell UTs, and versus transmit power of intra-cell UTs, respectively. It is observed that the MC-JCD-GAMP algorithm and the MC-JCD-GAMP-D algorithm have almost the same performance (which are the best over all), while the SCED algorithm has the worst performance in all cases. SCED algorithm estimates the channel from pilot symbols, while the JCED-Iterative and BiGAMP-based algorithms use the pilot symbols and also the estimated data symbols in order to estimate the channel. Moreover, JCED-Iterative and BiGAMP-based algorithms use the SCED algorithm as initialization, hence they have an obvious advantage compared to the SCED algorithm. In summary, joint channel and data estimation algorithms achieve better SER

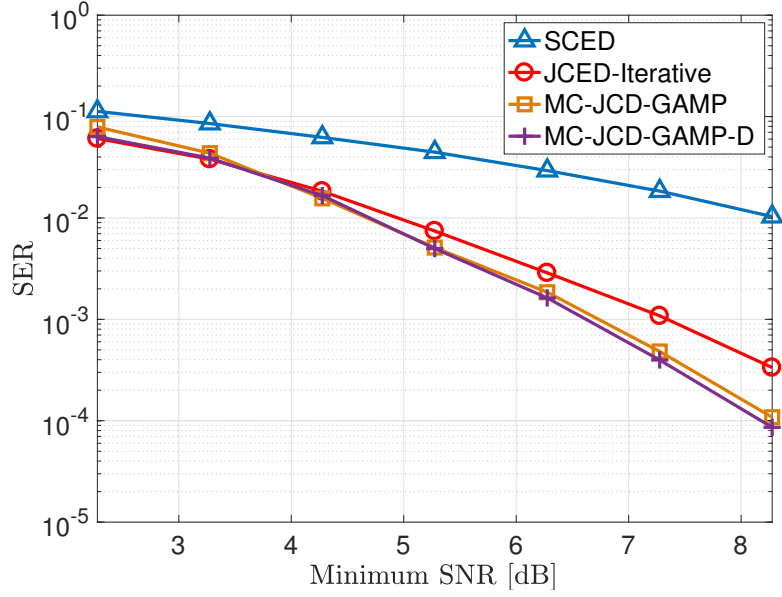


Figure 5.5: Symbol Error Rate versus minimum received SNR, for $N = 64$, $K = 10$ and $T_d = 240$.

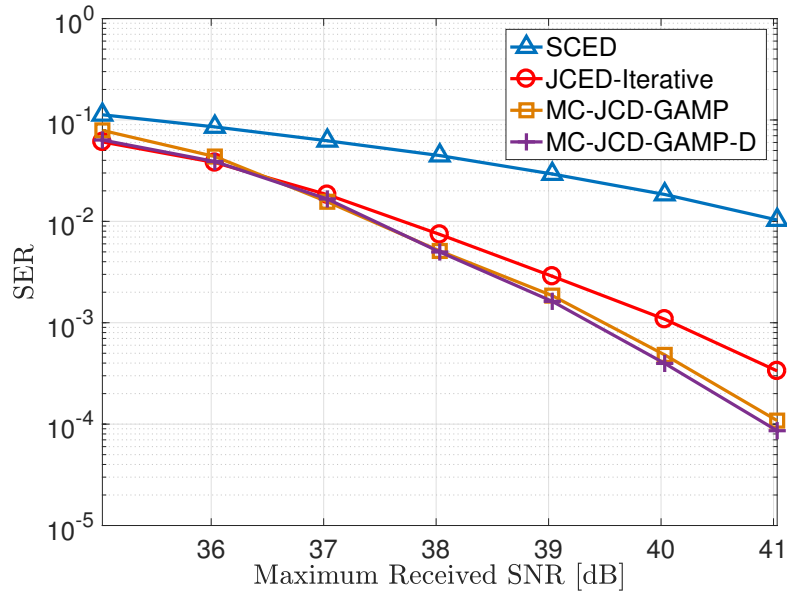


Figure 5.6: Symbol Error Rate versus maximum received SNR, for $N = 64$, $K = 10$ and $T_d = 240$.

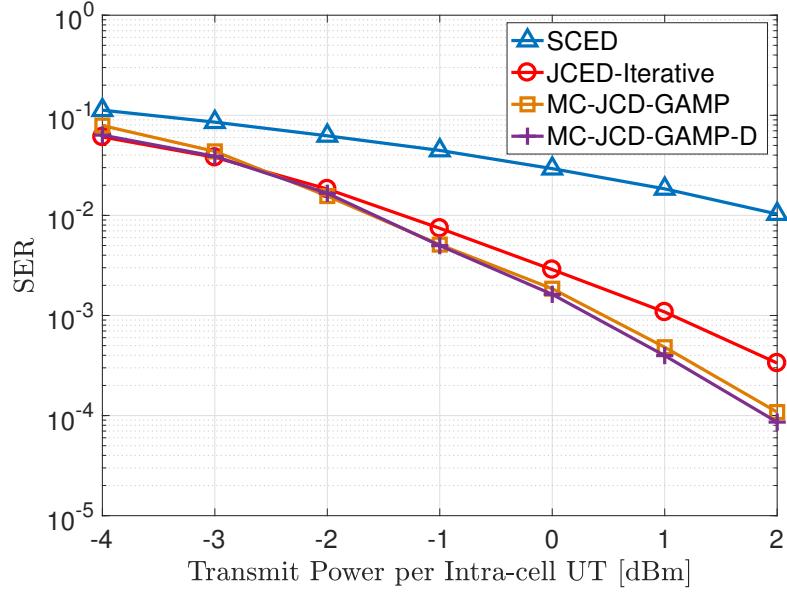


Figure 5.7: Symbol Error Rate versus transmit power per intra-cell UT, for $N = 64$, $K = 10$ and $T_d = 240$.

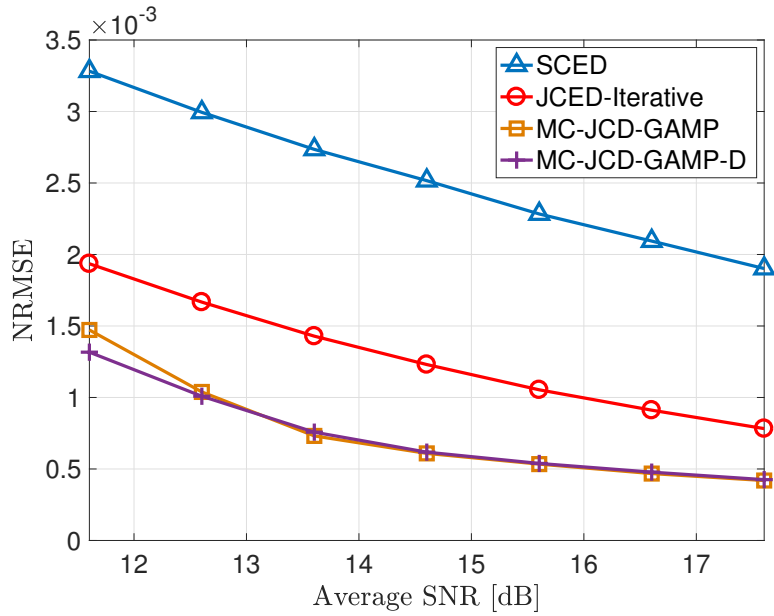


Figure 5.8: NRMSE versus average received SNR, for $N = 64$, $K = 10$ and $T_d = 240$.

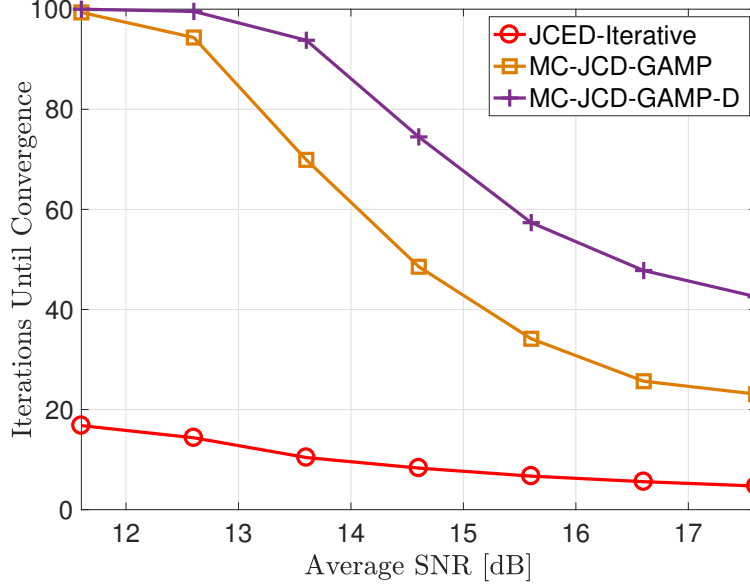


Figure 5.9: Iterations until convergence versus average received SNR, for $N = 64$, $K = 10$ and $T_d = 240$.

compared to the separate channel and data estimation algorithm.

Furthermore, as the transmit power of intra-cell UTs increases, the performance loss between SCED and JCED-Iterative algorithms and BiGAMP-based algorithms increases too. In more detail, at low SNRs, JCED-Iterative algorithm has slightly better SER than BiGAMP-based algorithm, while at mean and high SNRs the performance loss of JCED-Iterative is approximately from 0.1 dB to 1 dB compared to BiGAMP-based algorithms. In addition, it is observed that the performance loss of SCED algorithm is approximately from 2.4 dB to 3.6 dB.

In Fig. 5.8, NRMSE performances of SCED, JCED-Iterative, MC-JCD-GAMP and MC-JCD-GAMP-D algorithms are shown versus average received SNR at BS from intra-cell UTs. It is observed that the SCED algorithm estimates the intra-cell channel with the highest MSE compared to JCED-Iterative and BiGAMP-based algorithms. Furthermore, simulations show that BiGAMP-based algorithms achieve better estimation compared to JCED-Iterative algorithm. Through simulations, it is proved that joint

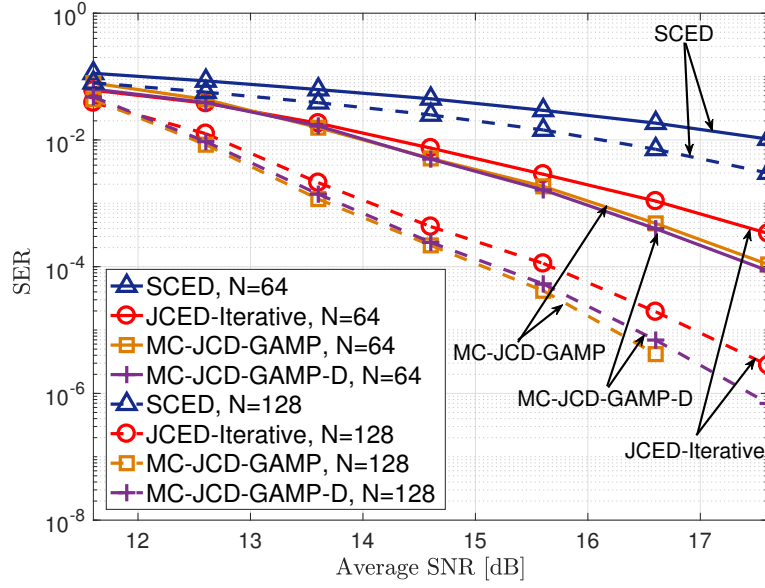


Figure 5.10: Symbol Error Rate versus average received SNR, for $N = 64$ or $N = 128$, $K = 10$ and $T_d = 240$.

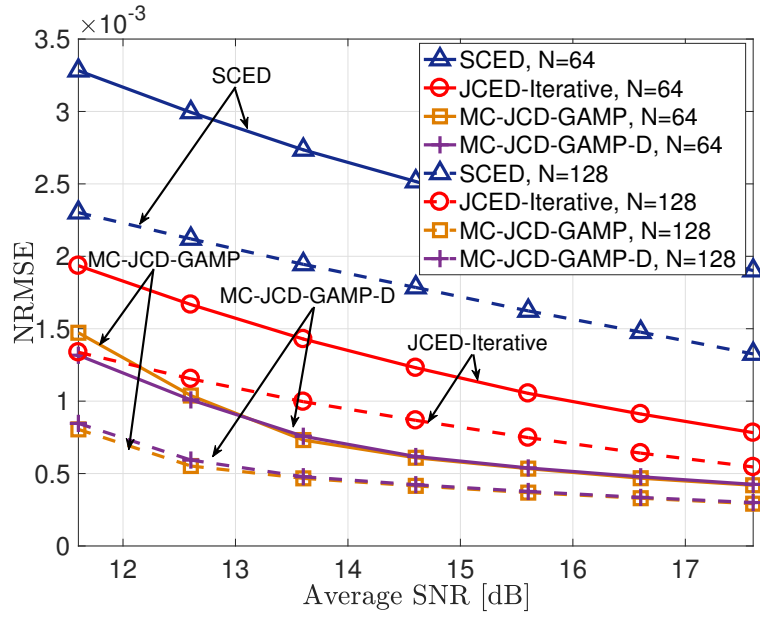


Figure 5.11: NRMSE versus average received SNR, for $N = 64$ or $N = 128$, $K = 10$ and $T_d = 240$.

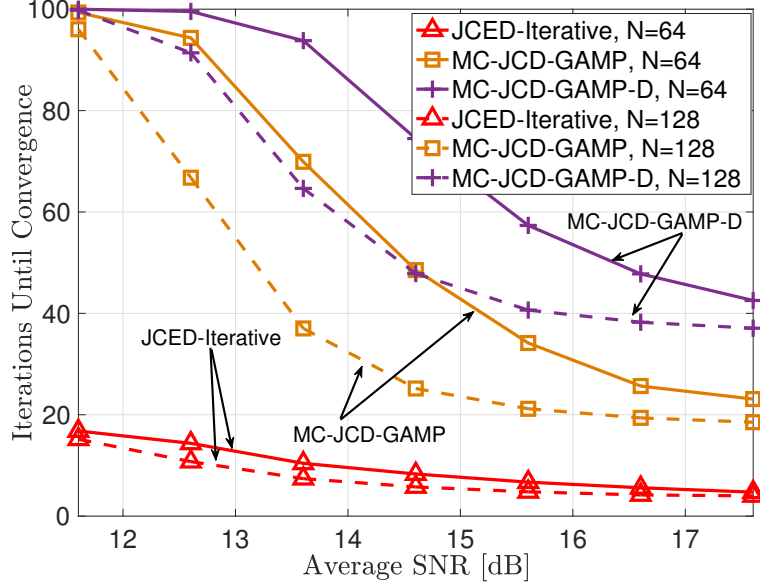


Figure 5.12: Iterations until convergence versus average received SNR, for $N = 64$ or $N = 128$, $K = 10$ and $T_d = 240$.

channel and data estimation algorithms achieve better performance (SER or NRSME) compared to separate channel and data estimation algorithms.

Fig. 5.9 depicts the number of iterations needed for BiGAMP-based algorithm and JCED algorithm to converge. It is shown that JCED algorithm converges quite faster compared to BiGAMP-based algorithms. Moreover, it is noted that MC-JCD-GAMP algorithm needs a smaller number of iterations than MC-JCD-GAMP-D algorithm. In addition, the impact of SNR on the convergence of the algorithms is shown. As SNR increases, the iterations needed in order for the algorithms to converge decreases.

Fig. 5.10, Fig. 5.11 and Fig. 5.12 depict the performance of channel and data estimation algorithms versus the number of antennas the BS has. Hence, the same scenario as above is assumed, both with $N = 64$ antennas and with $N = 128$ antennas. As expected, the performance (SER or NRMSE) of the algorithms when the BS has 128 antennas is quite better compared to the case where the BS has 64 antennas. It is noted that for BiGAMP-based algorithms with 128 received antennas at the BS, the performance loss

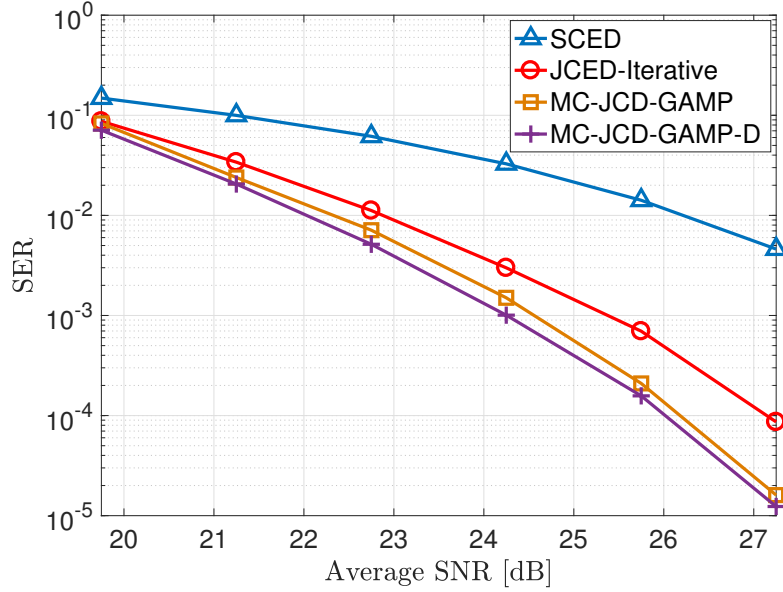


Figure 5.13: Symbol Error Rate versus average received SNR, for $N = 64$, $K = 15$ and $T_d = 360$.

(SER) is approximately between 1 dB and 2.6 dB compared to 64 received antennas at the BS. Furthermore, in Fig. 5.12 it is shown that algorithms converge faster in the case where the BS has 128 antennas, since better data estimation is provided when more received antennas are used.

5.3 Second Scenario (64 Received Antennas, 15 UTs per Cell)

The second set of simulations is conducted for $N = 64$ antennas at central BS, $K = 15$ UTs per cell and $T = 375$ symbols ($T_t = 15$ pilot symbols, $T_d = 350$ data symbols). In order to obtain different values for SNR and SINR, the transmit power of intra-cell UTs is varied from 5 dBm to 12.5 dBm, while the transmit power of inter-cell UTs is held constant at 12 dBm.

In Fig. 5.13, Fig. 5.14, SER performances of SCED, J-CED-Iterative, MC-JCD-GAMP, MC-JCD-GAMP-D algorithms are shown versus average received SNR at BS from intra-cell UTs, and versus transmit power of intra-

cell UTs, respectively. It is observed that MC-JCD-GAMP-D algorithm has slightly better performance compared to MC-JCD-GAMP algorithm (~ 0.15 dB performance loss). Performance loss of the JCED algorithm and the JCED-Iterative algorithm compared to MC-JCD-GAMP-D algorithms is approximately from 2.4 dB to 4.4 dB and from 0.01 dB to 1, respectively. Note that the maximum received SNR at BS from intra-cell UTs is 30.7 dB when transmit power of intra-cell UTs is 5 dBm, and 37.8 dB when transmit power of intra-cell UTs is 12.5 dBm. Moreover, the minimum received SNR at BS from intra-cell UTs is 13.5 dB when transmit power of intra-cell UTs is 5 dBm, and 21 dB when transmit power of intra-cell UTs is 12.5 dBm.

In Fig. 5.15, NRMSE performances of SCED, JCED-Iterative, MC-JCD-GAMP and MC-JCD-GAMP-D algorithms are shown versus average received SNR. It is observed again that the SCED algorithm estimates the intra-cell channels with the highest MSE compared to the JCED-Iterative and BiGAMP-based algorithms. Furthermore, simulations show that BiGAMP-based algorithms achieve better estimation compared to the JCED-Iterative

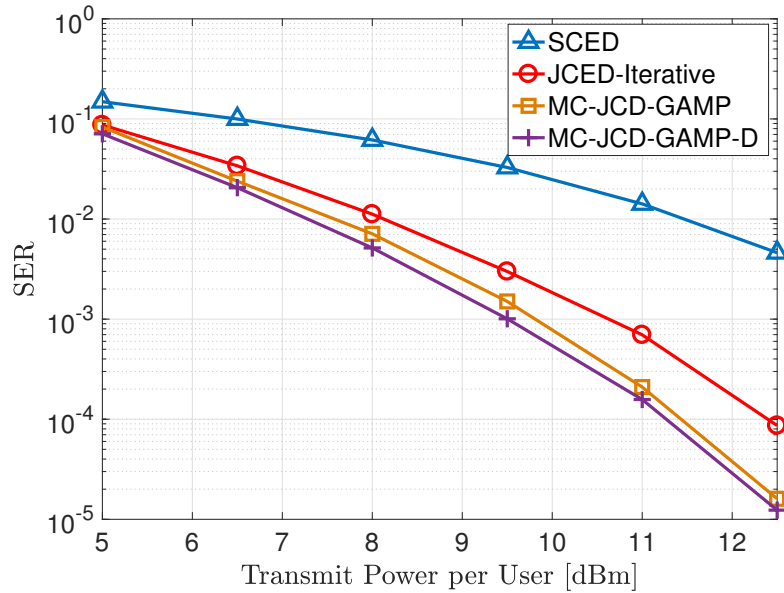


Figure 5.14: Symbol Error Rate versus transmit power per intra-cell UT, for $N = 64$, $K = 15$ and $T_d = 360$.

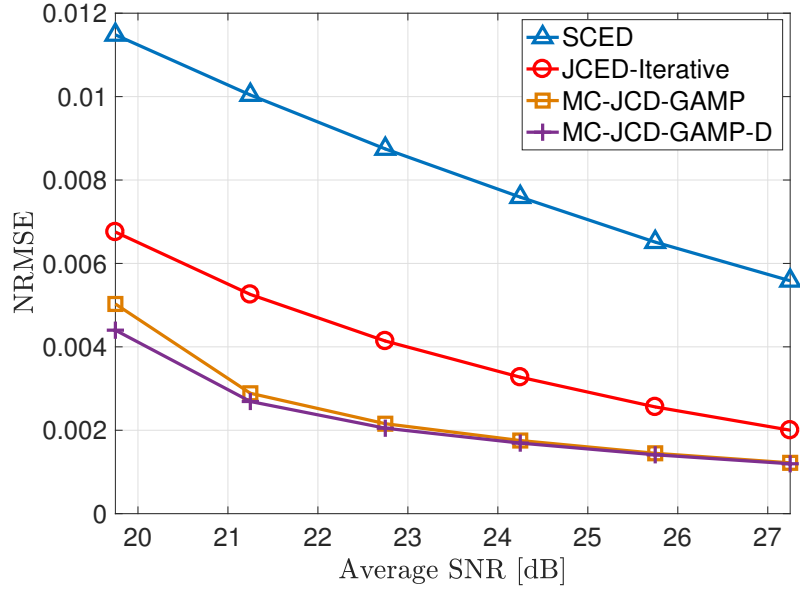


Figure 5.15: NRMSE versus average received SNR, for $N = 64$, $K = 15$ and $T_d = 360$.

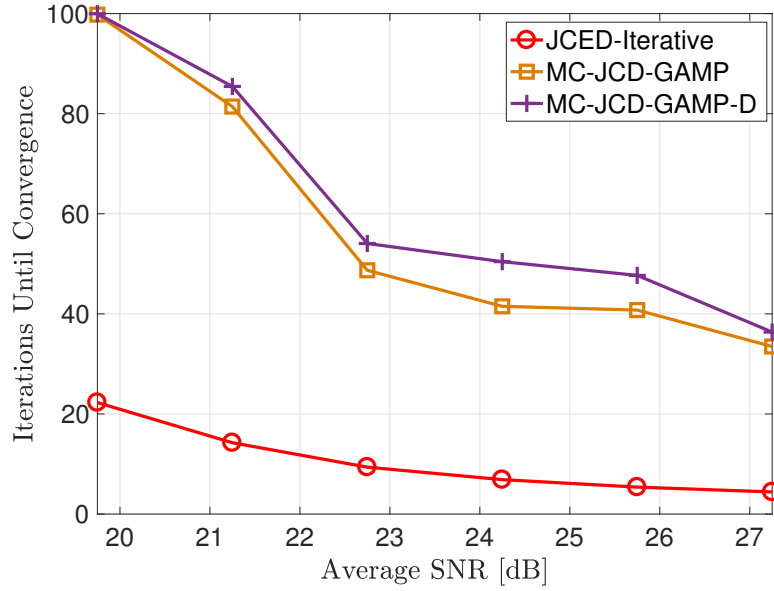


Figure 5.16: Iterations until convergence versus average received SNR, for $N = 64$, $K = 15$ and $T_d = 360$.

algorithm, while BiGAMP-based algorithms have almost the same performance.

Fig. 5.16, depicts the number of iterations needed in order for BiGAMP-based algorithm and JCED algorithm to converge. It is shown again, as in the first scenario Sec. 5.2, that the JCED algorithm converges quite faster compared to BiGAMP-based algorithms.

5.4 Third Scenario (Moving Intra-cell UT)

The third set of simulations is conducted for $N = 64$ antennas at central BS, $K = 15$ UTs per cell and $T = 375$ symbols ($T_t = 15$ pilot symbols, $T_d = 350$ data symbols). In this section, intra-cell UT k is placed at the edge of the cell and moves 50 m closed to BS at each simulation. Transmit power of intra-cell UT k is held constant at 3 dBm, while the transmit power of $K - 1$ intra-cell UTs and inter-cell UTs is held constant at 5 dBm. Fig. 5.17

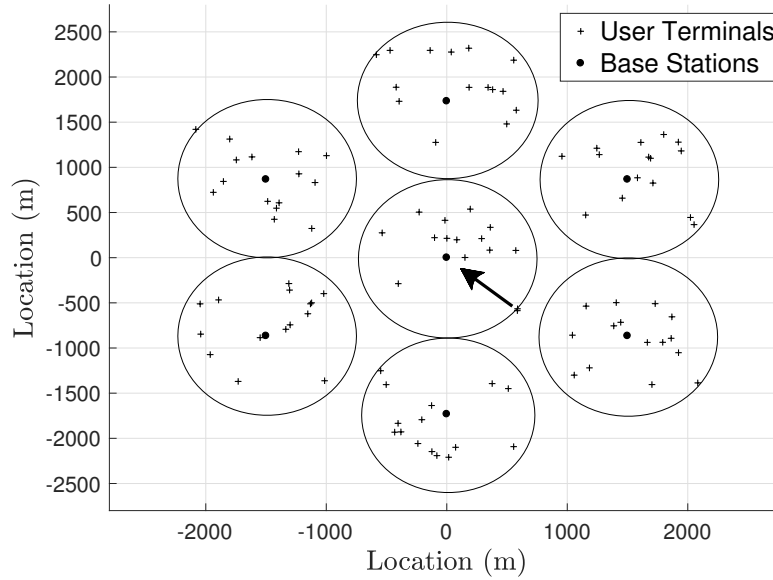


Figure 5.17: Topology of 7-cell hexagonal system with $K = 15$ UTs uniformly distributed around each BS, where arrow represents the movement of intra-cell UT k .

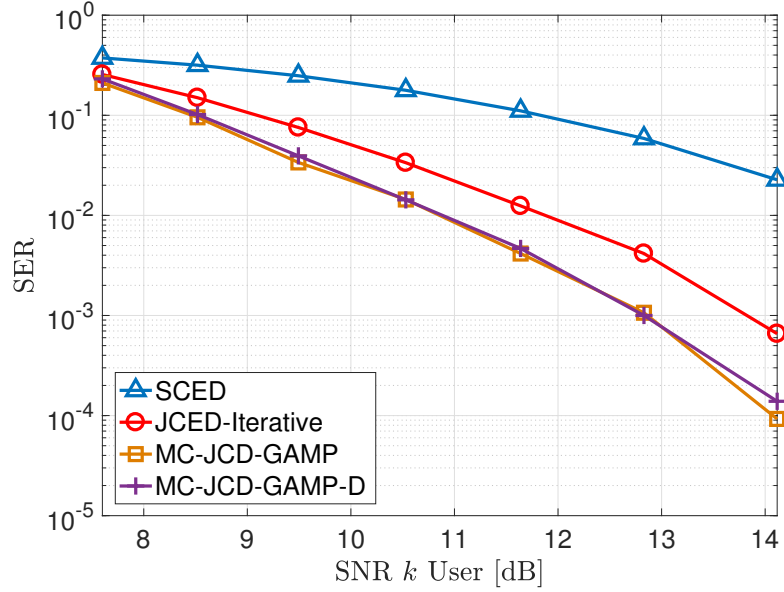


Figure 5.18: Symbol Error Rate versus received SNR from intra-cell UT k , for $N = 64$, $K = 15$ and $T_d = 360$.

represent the topology of scenario that studied.

In Fig. 5.18 and Fig. 5.19, SER performances of SCED, J-CED-Iterative, MC-JCD-GAMP, MC-JCD-GAMP-D algorithms are shown versus received SNR at BS from intra-cell UT k and versus the distance of intra-cell UT k from BS, respectively. It is observed again, that the MC-JCD-GAMP-D algorithm has almost the same performance with the MC-JCD-GAMP algorithm. Performance loss of the JCED algorithm and the JCED-Iterative algorithm compared to MC-JCD-GAMP-D algorithms is approximately from 1.5 dB to 4.1 dB and from 0.05 dB to 1.1, respectively. Also note that the maximum received SNR at BS from intra-cell UTs is 30.4 dB in all cases. In addition, the minimum received SNR at BS from intra-cell UTs is the received SNR of UT k when the distance of intra-cell UT k from BS is between 650 m and 900 m, and 13.8 dB in the remaining case (in this case the minimum received SNR emanates from UT m , where $m \neq k$).

In Fig. 5.20, NRMSE performances of SCED, JCED-Iterative, MC-JCD-GAMP and MC-JCD-GAMP-D algorithms are shown versus received SNR

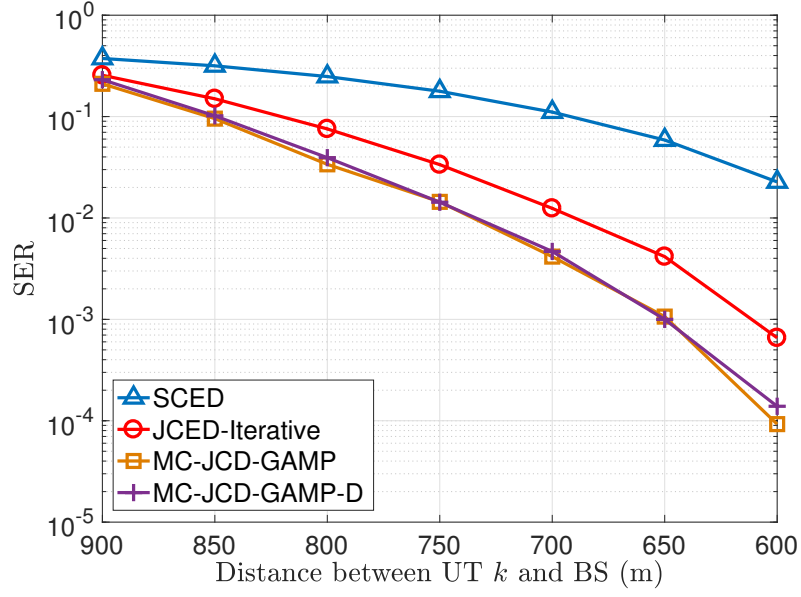


Figure 5.19: Symbol Error Rate versus distance of intra-cell UT k from BS, for $N = 64$, $K = 15$ and $T_d = 360$.

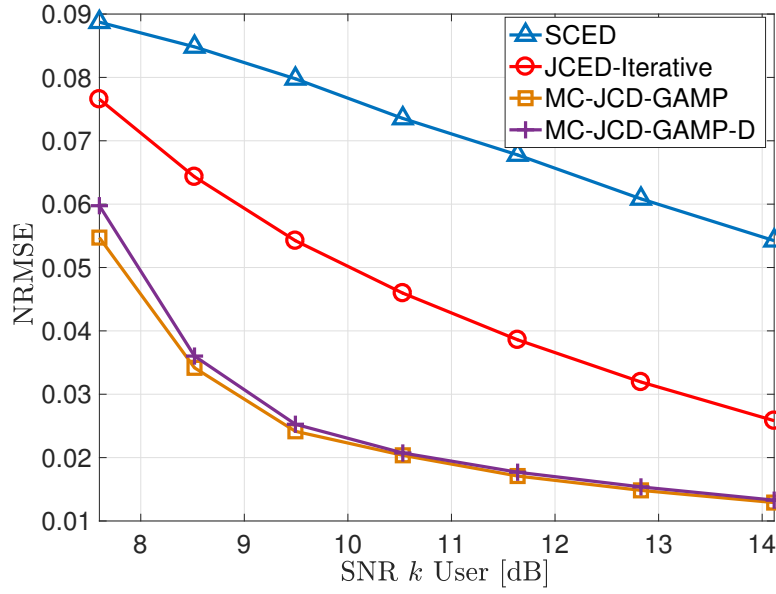


Figure 5.20: NRMSE versus received SNR from intra-cell UT k , for $N = 64$, $K = 15$ and $T_d = 360$.

from intra-cell UT k . It is noted again that the SCED algorithm estimates the intra-cell channels with the highest MSE compared to the JCED-Iterative and BiGAMP-based algorithms. Furthermore, simulations proved once more that BiGAMP-based algorithms have almost the same performance.

Fig. 5.21, depicts the number of iterations needed in order for BiGAMP-based algorithm and JCED algorithm to converge. It is shown, that the JCED algorithm converges quite faster compared to BiGAMP-based algorithms. Moreover MC-JCD-GAMP-D algorithm converges slower compared to the MC-JCD-GAMP algorithm at low SNRs, while at high SNRs the MC-JCD-GAMP-D converges faster compared to the MC-JCD-GAMP algorithm. This observation maybe can be explained by damping factor number that chosen in the MC-JCD-GAMP algorithm, which may be not the optimal value at low SNRs. As future work, damping factor and the above observation need further research in order to reach safer conclusions concerning these topics.

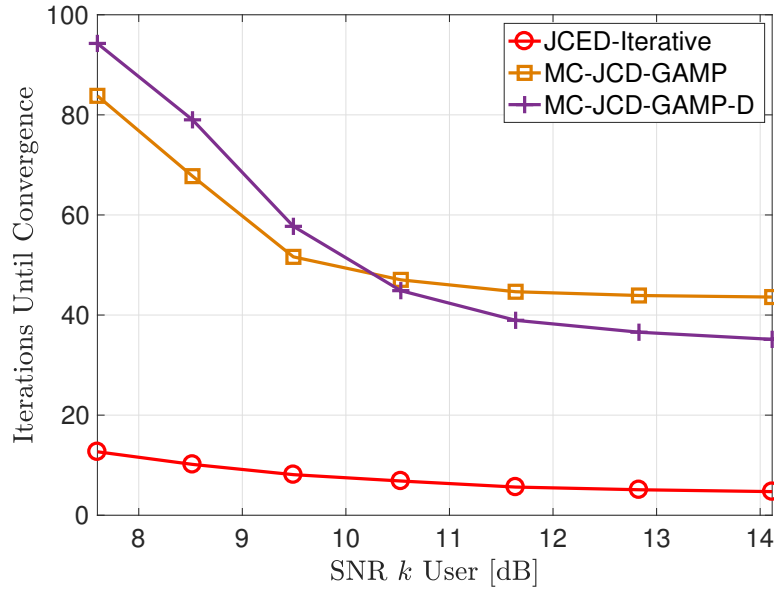


Figure 5.21: Iterations until convergence versus received SNR from intra-cell UT k , for $N = 64$, $K = 15$ and $T_d = 360$.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The main focus of this work was to implement the bilinear generalized approximate message passing algorithm (BiGAMP-based algorithm) in multi-cell uplink interference-limited massive MIMO networks. A general model is considered, which takes into account path-loss and pilot contamination from adjacent cell. The separate MMSE channel estimation and linear MMSE data detection (SCED) algorithm is presented. Moreover, based on the above algorithm, the heuristic iterative channel estimation and data detection (JCED-Iterative) algorithm was presented, which estimates the data and the channel in a joint way.

JCD-GAMP algorithm [12] is modified (MC-JCD-GAMP) in order to use it in multi-cell massive MIMO network. New initialization is proposed. Moreover, the MC-JCD-GAMP-D algorithm is proposed, in which pilot symbols are used only in initialization phase while the message passing procedure uses the data symbols. It is noted that the performance of BiGAMP-based algorithms strongly depends on initialization and on the damping factor. Hence, tuning is extremely important for the convergence of the above algorithms.

Simulation results under realistic assumptions and system dimensions show that BiGAMP-based algorithms achieve better performance compared to the SCED algorithm, as well as the JCED-Iterative algorithm. On the other hand, the second algorithm converges quite faster compared to BiGAMP-based algorithms in all cases. The idea using the estimated channel and detected data of SCED algorithm as input to BiGAMP-based algorithms strongly improves the performance of the above algorithms. Nevertheless,

the above idea increases the overall computation complexity. Note that BiGAMP-based algorithms have considerably smaller computation complexity compared to JCED algorithm. Therefore, there exist a numerous of interesting convergence-accuracy trade-offs between the above algorithms. Furthermore, it is observed that the performance of joint channel estimation and detection algorithms is better than the performance of separate channel estimation and detection algorithm.

In addition, it is shown that damping factor is an important parameter, which affects the overall performance of BiGAMP-based algorithms, especially at low SNRs where BiGAMP-based algorithms may be diverge. Also, it is shown that MC-JCD-GAMP and MC-JCD-GAMP algorithms have approximately the same performance (SER or NRMSE). In addition, it is noted that the second algorithm needs quite smaller damping factor in order to converge compared to the first algorithm, while it is also needs more iterations. On the other hand, the computation complexity of the MC-JCD-GAMP-D algorithm in message passing procedure is smaller compared to MC-JCD-GAMP algorithm.

Summary, BiGAMP-based algorithms, which have scalar operations with simple computation methods, and also JCED algorithm offer an efficient solution to joint channel and data estimation problem in multi-cell interference-limited massive MIMO network.

6.2 Future Work

The main focus of this work is to implement algorithms for joint channel and data estimation in unquantized multi-cell massive MIMO network. Hence, BiGAMP-based algorithms and JCED-Iterative can be implemented in quantized multi-cell massive MIMO network. Another concept is the use of different modulations and also different estimators in BiGAMP-based algorithms. Moreover, most problems that investigated in literature in multi-cell networks consider the interference channels as known at BSs, i.e. downlink beamforming problem. In practice, these channels are unknown. Hence, BiGAMP-based algorithms can be used in order to estimate both the chan-

nels between intra-cell UTs and BS and also the channels between inter-cell (interference) UTs and BS.

Appendix A

Mean and Variance of Noise in JCED-Iterative

From Eq. 3.12 is considered that

$$\mathbf{y}_{\text{est},j,k}^{\text{tr}} = \mathbf{Y}_j \mathbf{x}_{\text{MP},j,k} = \mathbf{h}_{jj,k} + \sum_{l \neq j} \mathbf{h}_{jl,k} + \bar{\mathbf{w}}_{j,k}. \quad (\text{A.1})$$

Hence,

$$\bar{\mathbf{w}}_{j,k} = \mathbf{W}_j \mathbf{x}_{\text{MP},j,k} \quad (\text{A.2})$$

with each element $W_{j,nt}$ of \mathbf{W}_j follows $\mathcal{N}_{\mathbb{C}}(W_{j,nt}; 0, \sigma_w^2)$. Moreover,

$$\mathbf{W}_j = \begin{bmatrix} \mathbf{w}_{j,1} \\ \mathbf{w}_{j,2} \\ \vdots \\ \mathbf{w}_{j,N} \end{bmatrix}, \quad (\text{A.3})$$

$\mathbf{w}_{j,n} \in \mathbb{C}^{1 \times T}$. Therefore,

$$\mathbb{E}[\bar{w}_{j,kn}] = \mathbb{E}[\mathbf{w}_{j,n} \mathbf{x}_{\text{MP},j,k}] = \mathbb{E}\left[\sum_{t=1}^T w_{j,nt} x_{\text{MP},j,tk}\right] \quad (\text{A.4})$$

$$= \sum_{t=1}^T \mathbb{E}[w_{j,nt} x_{\text{MP},j,tk}] \quad (\text{A.5})$$

$$= \sum_{t=1}^T x_{\text{MP},j,tk} \mathbb{E}[w_{j,nt}] = 0 \quad (\text{A.6})$$

and

$$\mathbb{E}[\bar{w}_{j,kn}\bar{w}_{j,kn}^*] = \mathbb{E}[\mathbf{w}_{j,n}\mathbf{x}_{\text{MP},j,k}(\mathbf{x}_{\text{MP},j,k})^H\mathbf{w}_{j,n}^H] \quad (\text{A.7})$$

$$= \mathbb{E}[\text{trace}(\mathbf{w}_{j,n}\mathbf{x}_{\text{MP},j,k}(\mathbf{x}_{\text{MP},j,k})^H\mathbf{w}_{j,n}^H)] \quad (\text{A.8})$$

$$= \mathbb{E}[\text{trace}(\mathbf{w}_{j,n}^H\mathbf{w}_{j,n}\mathbf{x}_{\text{MP},j,k}\mathbf{x}_{\text{MP},j,k}^H)] \quad (\text{A.9})$$

$$\stackrel{(a)}{=} \text{trace}(\mathbb{E}[\mathbf{w}_{j,n}^H\mathbf{w}_{j,n}\mathbf{x}_{\text{MP},j,k}\mathbf{x}_{\text{MP},j,k}^H]) \quad (\text{A.10})$$

$$= \text{trace}(\mathbb{E}[\mathbf{w}_{j,n}^H\mathbf{w}_{j,n}]\mathbf{x}_{\text{MP},j,k}\mathbf{x}_{\text{MP},j,k}^H) \quad (\text{A.11})$$

$$= \text{trace}(\sigma_w^2\mathbf{I}_T\mathbf{x}_{\text{MP},j,k}\mathbf{x}_{\text{MP},j,k}^H) \quad (\text{A.12})$$

$$= \text{trace}(\mathbf{x}_{\text{MP},j,k}^H\sigma_w^2\mathbf{I}_T\mathbf{x}_{\text{MP},j,k}) \quad (\text{A.13})$$

$$= \|\mathbf{x}_{\text{MP},j,k}\|_2^2\sigma_w^2 \quad (\text{A.14})$$

where in (a) we used the linearity of trace and expectation operators.

In summary, $\bar{\mathbf{w}}_{j,k} \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{w}}_{j,k}; \mathbf{0}, \|\mathbf{x}_{\text{MP},j,k}\|_2^2\sigma_w^2\mathbf{I}_N)$.

Appendix B

Mean and Variance of Interference Signal

Consider a SISO version of system Eq. 2.1, $Y_j = Z_j + E_j + W_j$, (for $N = 1, T = 1$) where $Z_j = \mathbf{h}_{jj}\mathbf{x}_j$, $E_j = \sum_{l \neq j}^L \mathbf{h}_{jl}\mathbf{x}_l$, with $\mathbf{h}_{jj} \in \mathbb{C}^{1 \times K}$, $\mathbf{h}_{jl} \in \mathbb{C}^{1 \times K}$, $\mathbf{x}_j \in \mathbb{C}^K$ and $\mathbf{x}_l \in \mathbb{C}^K$. Furthermore, $W_j \sim \mathcal{N}_{\mathbb{C}}(W_j; 0, \sigma_w^2)$. QPSK modulations is assumed, hence all data symbols has the same probability which is equal to $1/4$. In more detail,

$$\mathbf{x}_l = \begin{bmatrix} x_{l,1} \\ x_{l,2} \\ \vdots \\ x_{l,k} \end{bmatrix} \quad (\text{B.1})$$

and

$$\mathbf{h}_{jl} = [h_{jl,1} \ h_{jl,2} \ \dots \ h_{jl,k}] \quad (\text{B.2})$$

with $h_{jl,k} \sim \mathcal{N}_{\mathbb{C}}(h_{jl,k}; 0, \sigma_{h_{jl,k}}^2)$. Therefore, for each UT k

$$\mathbb{E}[h_{jl,k}x_{l,k}] = \mathbb{E}[h_{jl,k}]\mathbb{E}[x_{l,k}] = 0 \quad (\text{B.3})$$

and

$$\sigma_{h_{jl,k}x_{l,k}}^2 = \mathbb{E}[h_{jl,k}x_{l,k}(h_{jl,k}x_{l,k})^*] = \mathbb{E}[h_{jl,k}x_{l,k}x_{l,k}^*h_{jl,k}^*] \quad (\text{B.4})$$

$$= \mathbb{E}[h_{jl,k}|x_{l,k}|^2h_{jl,k}^*] \quad (\text{B.5})$$

$$= \mathbb{E}[|x_{l,k}|^2]\mathbb{E}[h_{jl,k}h_{jl,k}^*] \quad (\text{B.6})$$

$$= \sigma_{h_{jl,k}}^2. \quad (\text{B.7})$$

Hence, for all UTs in l -th cell

$$\mathbb{E}[\mathbf{h}_{jl}\mathbf{x}_l] = \mathbb{E}\left[\sum_{k=1}^K h_{jl,k}x_{l,k}\right] \quad (\text{B.8})$$

$$= \sum_{k=1}^K \mathbb{E}[h_{jl,k}x_{l,k}] \quad (\text{B.9})$$

$$= 0. \quad (\text{B.10})$$

Since the elements of vector \mathbf{x}_l are i.i.d. and belong to a QPSK constellation it follows that $\mathbb{E}[\mathbf{x}_l\mathbf{x}_l^H] = \mathbf{I}_K$, thus

$$\sigma_{\mathbf{h}_{jl}\mathbf{x}_l}^2 = \mathbb{E}[\mathbf{h}_{jl}\mathbf{x}_l(\mathbf{h}_{jl}\mathbf{x}_l)^*] = \mathbb{E}[\mathbf{h}_{jl}\mathbf{x}_l\mathbf{x}_l^H\mathbf{h}_{jl}^H] \quad (\text{B.11})$$

$$= \mathbb{E}[\text{trace}(\mathbf{h}_{jl}\mathbf{x}_l\mathbf{x}_l^H\mathbf{h}_{jl}^H)] \quad (\text{B.12})$$

$$= \mathbb{E}[\text{trace}(\mathbf{h}_{jl}^H\mathbf{h}_{jl}\mathbf{x}_l\mathbf{x}_l^H)] \quad (\text{B.13})$$

$$\stackrel{(a)}{=} \text{trace}(\mathbb{E}[\mathbf{h}_{jl}^H\mathbf{h}_{jl}\mathbf{x}_l\mathbf{x}_l^H]) \quad (\text{B.14})$$

$$\stackrel{(b)}{=} \text{trace}(\mathbb{E}[\mathbf{h}_{jl}^H\mathbf{h}_{jl}] \mathbb{E}[\mathbf{x}_l\mathbf{x}_l^H]) \quad (\text{B.15})$$

$$= \text{trace}(\mathbb{E}[\mathbf{h}_{jl}^H\mathbf{h}_{jl}] \mathbf{I}_K) \stackrel{(c)}{=} \mathbb{E}[\text{trace}(\mathbf{h}_{jl}^H\mathbf{h}_{jl})] \quad (\text{B.16})$$

$$= \mathbb{E}[\mathbf{h}_{jl}\mathbf{h}_{jl}^H] = \mathbb{E}\left[\sum_{k=1}^K h_{jl,k}h_{jl,k}^*\right] \quad (\text{B.17})$$

$$= \sum_{k=1}^K \mathbb{E}[h_{jl,k}h_{jl,k}^*] \quad (\text{B.18})$$

$$= \sum_{k=1}^K \sigma_{h_{jl,k}}^2. \quad (\text{B.19})$$

where in (a) and (c) we used the linearity of trace and expectation operators and in (b) we used the independence of \mathbf{h}_{jl} and \mathbf{x}_l .

Finally, for $E_j = \sum_{l \neq j}^L \mathbf{h}_{jl} \mathbf{x}_l$

$$\mathbb{E}[E_j] = \mathbb{E} \left[\sum_{l=1}^L \mathbf{h}_{jl} \mathbf{x}_l \right] \quad (\text{B.20})$$

$$= \sum_{l=1}^L \mathbb{E}[\mathbf{h}_{jl} \mathbf{x}_l] \quad (\text{B.21})$$

$$= 0 \quad (\text{B.22})$$

and

$$\sigma_{E_j}^2 = \mathbb{E} \left[\sum_{l \neq j} \mathbf{h}_{jl} \mathbf{x}_l \left(\sum_{l \neq j} \mathbf{h}_{jl} \mathbf{x}_l \right)^* \right] = \mathbb{E} \left[\sum_{l \neq j} \sum_{l' \neq j} \mathbf{h}_{jl} \mathbf{x}_l \left(\mathbf{h}_{jl'} \mathbf{x}_{l'} \right)^* \right] \quad (\text{B.23})$$

$$= \mathbb{E} \left[\sum_{l \neq j} \mathbf{h}_{jl} \mathbf{x}_l \left(\mathbf{h}_{jl} \mathbf{x}_l \right)^* \right] \quad (\text{B.24})$$

$$= \sum_{l \neq j} \mathbb{E} \left[\mathbf{h}_{jl} \mathbf{x}_l \left(\mathbf{h}_{jl} \mathbf{x}_l \right)^* \right] \quad (\text{B.25})$$

$$= \sum_{l \neq j} \sum_{k=1}^K \sigma_{h_{jl,k}}^2. \quad (\text{B.26})$$

In summary, $E_j \sim \mathcal{N}_{\mathbb{C}}(E_j; 0, \sum_{l \neq j} \sum_{k=1}^K \sigma_{h_{jl,k}}^2)$.

Appendix C

Product of Two Gaussian Distributions

In [22, (A.7)] it is analyzed that the product two Gaussian distributions gives another Gaussian distribution. If $\mathcal{N}_{\mathbb{C}}(x; a, A)$ and $\mathcal{N}_{\mathbb{C}}(x; b, B)$

$$\mathcal{N}_{\mathbb{C}}(x; a, A)\mathcal{N}_{\mathbb{C}}(x; b, B) = \mathcal{D}\mathcal{N}_{\mathbb{C}}(x; c, C) \quad (\text{C.1})$$

where $c = C(A^{-1}a + B^{-1}b)$, $C = (A^{-1} + B^{-1})^{-1}$, and $\mathcal{D} = \mathcal{N}_{\mathbb{C}}(0; a - b, A + B)$. Hence, for Eq. 4.6

$$\begin{aligned} f_{Y|Z}(Y|Z)f_Z(Z) &= \mathcal{N}_{\mathbb{C}}(Z; Y, \sigma_u^2)\mathcal{N}_{\mathbb{C}}(Z; \hat{p}, v^p) \\ &= \mathcal{D} \cdot \mathcal{N}_{\mathbb{C}}(Z; c, C). \end{aligned} \quad (\text{C.2})$$

with

$$C = \left(\frac{1}{\sigma_u^2} + \frac{1}{v^p} \right)^{-1} = \frac{\sigma_u^2 v^p}{v^p + \sigma_u^2} \quad (\text{C.3})$$

and

$$c = \frac{\sigma_u^2 v^p}{v^p + \sigma_u^2} \left(\frac{Y}{\sigma_u^2} + \frac{\hat{p}}{v^p} \right) \quad (\text{C.4})$$

$$= C \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 v^p} \quad (\text{C.5})$$

$$= \frac{\sigma_u^2 v^p}{v^p + \sigma_u^2} \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 v^p} \quad (\text{C.6})$$

$$= \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 + v^p}. \quad (\text{C.7})$$

Also,

$$\mathcal{D} = \mathcal{N}_{\mathbb{C}}(0; Y - \hat{p}, \sigma_u^2 + v^p). \quad (\text{C.8})$$

Therefore,

$$f_{Y|Z}(Y|Z) = \mathcal{N}_{\mathbb{C}}\left(Z; \frac{v^p Y + \sigma_u^2 \hat{p}}{\sigma_u^2 + v^p}, \frac{\sigma_u^2 v^p}{\sigma_u^2 + v^p}\right). \quad (\text{C.9})$$

Appendix D

Estimator for QPSK Data Symbols

If QPSK modulation is considered, 2×2 points are selected by

$$\mathcal{B} = \frac{1}{\sqrt{2}} \times \{1 + j, -1 + j, 1 - j, -1 - j\} \quad (\text{D.1})$$

and, $\Pr(X_d) = 1/4$ for $X_d \in \mathcal{B}$. Conditional mean of X_d is

$$\hat{X}_d = \sum_{X_d \in \mathcal{B}} X_d \mathcal{P}(X_d) \quad (\text{D.2})$$

where the posterior probability mass function (pmf) $\mathcal{P}(X_d)$ is defined as

$$\mathcal{P}(X_d) = \frac{\mathcal{N}_{\mathbb{C}}(X_d; \hat{r}, u^r) \Pr(X_d)}{\sum_{X'_d} \mathcal{N}_{\mathbb{C}}(X'_d; \hat{r}, u^r) \Pr(X'_d)}. \quad (\text{D.3})$$

Hence,

$$\mathcal{P}(X_d) = \frac{\mathcal{N}_{\mathbb{C}}(X_d; \hat{r}, u^r)^{\frac{1}{4}}}{\frac{1}{4\pi u^r} \left[e^{-|\frac{1+j}{\sqrt{2}} - \hat{r}|^2 / u^r} + e^{-|\frac{1-j}{\sqrt{2}} - \hat{r}|^2 / u^r} + e^{-|\frac{-1+j}{\sqrt{2}} - \hat{r}|^2 / u^r} + e^{-|\frac{-1-j}{\sqrt{2}} - \hat{r}|^2 / u^r} \right]}. \quad (\text{D.4})$$

Therefore,

$$\hat{X}_d = \frac{\frac{1+j}{\sqrt{2}} e^{-|\frac{1+j}{\sqrt{2}} - \hat{r}|^2 / u^r} + \frac{1-j}{\sqrt{2}} e^{-|\frac{1-j}{\sqrt{2}} - \hat{r}|^2 / u^r} + \frac{-1+j}{\sqrt{2}} e^{-|\frac{-1+j}{\sqrt{2}} - \hat{r}|^2 / u^r} + \frac{-1-j}{\sqrt{2}} e^{-|\frac{-1-j}{\sqrt{2}} - \hat{r}|^2 / u^r}}{e^{-|\frac{1+j}{\sqrt{2}} - \hat{r}|^2 / u^r} + e^{-|\frac{1-j}{\sqrt{2}} - \hat{r}|^2 / u^r} + e^{-|\frac{-1+j}{\sqrt{2}} - \hat{r}|^2 / u^r} + e^{-|\frac{-1-j}{\sqrt{2}} - \hat{r}|^2 / u^r}}. \quad (\text{D.5})$$

Set

$$\left| \frac{1+j}{\sqrt{2}} - \text{Re}(\hat{r}) - j\text{Im}(\hat{r}) \right|^2 \quad (\text{D.6})$$

$$= \left(\frac{1}{\sqrt{2}} - \text{Re}(\hat{r}) \right)^2 + \left(\frac{1}{\sqrt{2}} - \text{Im}(\hat{r}) \right)^2 \quad (\text{D.7})$$

$$= \frac{1}{2} + \text{Re}(\hat{r})^2 - \frac{2}{\sqrt{2}}\text{Re}(\hat{r}) + \frac{1}{2} + \text{Im}(\hat{r})^2 - \frac{2}{\sqrt{2}}\text{Im}(\hat{r}) \quad (\text{D.8})$$

Then, it is observed that $\frac{1}{2} + \text{Re}(\hat{r})^2$ and $\frac{1}{2} + \text{Im}(\hat{r})^2$ terms can be erased from Eq. D.5. If we define,

$$A = \frac{1+j}{\sqrt{2}} e^{\frac{-2 - \text{Re}(\hat{r}) - \text{Im}(\hat{r})}{u^r}}, \quad (\text{D.9})$$

$$B = \frac{1-j}{\sqrt{2}} e^{\frac{-2 - \text{Re}(\hat{r}) + \text{Im}(\hat{r})}{u^r}}, \quad (\text{D.10})$$

$$C = \frac{-1+j}{\sqrt{2}} e^{\frac{-2 + \text{Re}(\hat{r}) - \text{Im}(\hat{r})}{u^r}}, \quad (\text{D.11})$$

$$D = \frac{-1-j}{\sqrt{2}} e^{\frac{-2 + \text{Re}(\hat{r}) + \text{Im}(\hat{r})}{u^r}}, \quad (\text{D.12})$$

can be calculated that,

$$\hat{X}_d = \frac{\frac{1+j}{\sqrt{2}}A + \frac{1-j}{\sqrt{2}}B + \frac{-1+j}{\sqrt{2}}C + \frac{-1-j}{\sqrt{2}}D}{A + B + C + D} \quad (\text{D.13})$$

$$= \frac{(A + B - C - D) + j(A - B + C - D)}{\sqrt{2}(A + B + C + D)}. \quad (\text{D.14})$$

Then,

$$\frac{A + B - C - D}{\sqrt{2}(A + B + C + D)} \quad (\text{D.15})$$

$$= \frac{1}{\sqrt{2}} \frac{(e^{\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}} - e^{-\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}})(e^{\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}} + e^{-\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}})}{A + B + C + D} \quad (\text{D.16})$$

$$= \frac{1}{\sqrt{2}} \frac{4\sinh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}\right)\cosh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}\right)}{4\cosh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}\right)\cosh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}\right)} \quad (\text{D.17})$$

$$= \frac{1}{\sqrt{2}} \tanh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}\right), \quad (\text{D.18})$$

and

$$\frac{A - B + C - D}{\sqrt{2}(A + B + C + D)} \quad (\text{D.19})$$

$$= \frac{1}{\sqrt{2}} \frac{4\sinh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}\right)\cosh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}\right)}{4\cosh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}\right)\cosh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}\right)} \quad (\text{D.20})$$

$$= \frac{1}{\sqrt{2}} \tanh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}\right). \quad (\text{D.21})$$

Therefore,

$$\hat{X}_d = \frac{1}{\sqrt{2}} \tanh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}u^r}\right) + j \frac{1}{\sqrt{2}} \tanh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}u^r}\right). \quad (\text{D.22})$$

Moreover $u^x = \mathbb{E}[|X_d - \hat{X}_d|^2] = \sum_{X_d \in \mathcal{B}} |X_d - \hat{X}_d|^2 \mathcal{P}(X_d)$. Similarly can be calculated that

$$v^x = \frac{1}{2} \tanh\left(\frac{2\text{Re}(\hat{r})}{\sqrt{2}v^r}\right) + \frac{1}{2} \tanh\left(\frac{2\text{Im}(\hat{r})}{\sqrt{2}v^r}\right) - |\hat{X}_d|^2. \quad (\text{D.23})$$

Bibliography

- [1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, “Scenarios for 5G mobile and wireless communications: the vision of the metis project,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.
- [2] C. X. Wang, F. Haider, X. Gao, X. H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, “Cellular architecture and key technologies for 5G wireless communication networks,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.
- [3] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [5] E. Björnson, E. G. Larsson, and T. L. Marzetta, “Massive MIMO: ten myths and one critical question,” *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.
- [6] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, “An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications,” *Science China Information Sciences*, vol. 59, no. 8, Jun. 2016.

-
- [7] J. Hoydis, S. T. Brink, and M. Debbah, “Massive MIMO in the ul/dl of cellular networks: How many antennas do we need?” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
 - [8] J. Ma and L. Ping, “Data-aided channel estimation in large antenna systems,” in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Jun. 2014, pp. 4626–4631.
 - [9] K. Takeuchi, R. R. Muller, M. Vehkaperä, and T. Tanaka, “On an achievable rate of large rayleigh block-fading MIMO channels with no csi,” *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.
 - [10] M. Jiang, J. Akhtman, and L. Hanzo, “Iterative joint channel estimation and multi-user detection for multiple-antenna aided OFDM systems,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2904–2914, Aug. 2007.
 - [11] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, “Pilot embedding for joint channel estimation and data detection in MIMO communication systems,” *IEEE Trans. Commun.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.
 - [12] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, “Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision adcs,” *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
 - [13] J. T. Parker, P. Schniter, and V. Cevher, “Bilinear generalized approximate message passing - part i: Derivation,” *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.
 - [14] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theor.*, vol. 47, pp. 498–519, Feb. 2001.
 - [15] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

-
- [16] D. Chu, “Polyphase codes with good periodic correlation properties,” *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.
 - [17] J. Mo, P. Schniter, and R. W. H. Jr., “Channel estimation in broadband millimeter wave MIMO systems with few-bit adcs,” *CoRR*, vol. abs/1610.02735, 2016. [Online]. Available: <http://arxiv.org/abs/1610.02735>
 - [18] *Physical Channels and Modulation*, 2015, 3GPP TS 36.211 V12.7.0 Std.
 - [19] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Massive MIMO performance evaluation based on measured propagation data,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.
 - [20] A. Albert, *Regression and the Moore-Penrose pseudoinverse*, ser. Mathematics in Science and Engineering. Burlington, MA: Elsevier, 1972. [Online]. Available: <http://cds.cern.ch/record/1253778>
 - [21] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” *IEEE Trans. Inf. Theory*, pp. 2168–2172, Jul. 2011.
 - [22] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.