MASSIVE MULTIPLE-INPUT MULTIPLE-OUTPUT COMMUNICATION SYSTEMS WITH LOW-RESOLUTION QUANTIZERS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ALİ BULUT ÜÇÜNCÜ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

DECEMBER 2021

Approval of the thesis:

MASSIVE MULTIPLE-INPUT MULTIPLE-OUTPUT COMMUNICATION SYSTEMS WITH LOW-RESOLUTION QUANTIZERS

submitted by **ALİ BULUT ÜÇÜNCÜ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** _____

Prof. Dr. Ali Özgür Yılmaz
Supervisor, **Electrical and Electronics Eng. Dept., METU** _____

Examining Committee Members:

Prof. Dr. Tolga Mete Duman
Electrical and Electronics Eng. Dept., Bilkent Univ. _____

Prof. Dr. Ali Özgür Yılmaz
Electrical and Electronics Eng. Dept., METU _____

Prof. Dr. Cenk Toker
Electrical and Electronics Eng. Dept., Hacettepe Univ. _____

Assoc. Prof. Dr. Ayşe Melda Yüksel Turgut
Electrical and Electronics Eng. Dept., METU _____

Assist. Prof. Dr. Gökhan Muzaffer Güvensen
Electrical and Electronics Eng. Dept., METU _____

Date: 15.12.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Ali Bulut Üçüncü

Signature :

ABSTRACT

MASSIVE MULTIPLE-INPUT MULTIPLE-OUTPUT COMMUNICATION SYSTEMS WITH LOW-RESOLUTION QUANTIZERS

Ali Bulut Üçüncü,
Ph.D., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Ali Özgür Yılmaz

December 2021, 161 pages

Low resolution analog-to-digital converters (ADC) attracted much attention for their use in massive multiple-input multiple-output (MIMO) systems due to their low power consumption and cost. In this thesis, we question whether large number of antennas present in massive MIMO is sufficient to provide an ultimate performance or additional sampling in time (temporal oversampling) will provide significant performance advantages. To begin with, we illustrate the benefits of oversampling in time for uplink massive MIMO systems with low-resolution ADCs in terms of symbol error rate (SER) and achievable rate for both single-carrier (SC) and multi-carrier modulation scenarios by deriving analytical bounds and with simulations. We also propose a sequantial linear minimum-mean-square error (LMMSE) based receiver as a low-complexity detector, which is much more feasible to implement compared to the zero-forcing (ZF) detector for temporally oversampled massive MIMO systems. We also examine and illustrate the benefits of temporal oversampling for quantized massive MIMO systems under adjacent channel interference caused by the non-linearity of the quantizers. According to the results that we obtain, it seems that temporal oversampling can be very beneficial and should always be considered for use in

quantized massive MIMO. Finally, we examine whether a low-complexity MIMO detector, which can outperform the existing detectors of similar complexity, can be proposed even without resorting temporal oversampling. For that purpose, we propose a near optimal factor-graph based Ungerboeck type detector with bi-directional decision feedback, along with the derivation of LMMSE channel estimator for quantized wideband SC-MIMO systems. The proposed detector is shown to outperform a representative detector of comparable complexity from the literature in terms of SER and achievable rate per user performance metrics.

Keywords: massive MIMO, analog-to-digital converter (ADC), quantization, one-bit, 1-bit, oversampling, performance analysis, channel estimation, low-resolution, uplink, flat fading, wideband, frequency-selective fading, single-carrier, OFDM, Ungerboeck, Bussgang decomposition, iterative detector, reduced-state, decision feedback, zero-forcing, maximal ratio combining, sequantial LMMSE.

ÖZ

DÜŞÜK ÇÖZÜNÜRLÜKLÜ NİCEMLEYİCİLERE SAHİP KİTLESEL ÇOK-GİRDİLİ ÇOK-ÇIKTILI HABERLEŞME SİSTEMLERİ

Ali Bulut Üçüncü,
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Ali Özgür Yılmaz

Aralık 2021 , 161 sayfa

Kitlesel çoklu-girdili çoklu-çıktılı (MIMO) sistemlerinde düşük çözünürlüğe sahip analogdan-sayısala-dönüştürücülerin (ADC) kullanımı sahip oldukları düşük maliyet ve güç tüketimlerinden ötürü ilgi uyandırmıştır. Bu tezde düşük çözünürlükteki ADC içeren kitlesel MIMO sistemlerinde var olan çok sayıda antene ek olarak zamanda aşırı örnekleme kullanılmasının önemli ölçüde avantaj sağlayıp sağlayamadığı sorgulanmaktadır. Başlangıç olarak tekli-taşıyıcılı ve bir-bitlik ADC kullanılan durumda aşırı örnekleme yönteminin SER ve erişilebilir oran başarımları açısından ciddi faydalar sağladığı hem analitik olarak elde edilen ifadeler ile hem de yapılan benzetimlerle gösterilmektedir. Bunun yanında kitlesel MIMO yapıları için aşırı örnekleme ile çalışan sıfır-zorlayıcı (ZF) alıcıya nazaran çok daha düşük karmaşıklığa sahip sıralı doğrusal en küçük karesel hata (LMMSE) tabanlı bir alıcı da önerilmektedir. Ayrıca nicemlemeli kitlesel MIMO yapıları için nicemleyicilerdeki doğrusalsızlık sonucunda ortaya çıkan yan bant girişiminin de aşırı örnekleme yöntemi ile bastırılabileceği yine bu çalışmada irdelenmektedir. Elde edilen bulgular, aşırı örnekleme tekniğinin nicemleme altında çalışan kitlesel MIMO yapılarında kullanımının her zaman değerlendi-

rilmesi gerektiğini göstermektedir. Son olarak nicemleme altındaki kitlesel MIMO yapıları için aşırı örnekleme tekniği kullanılmadığı durumda literatürde var olan alıcılardan daha iyi başarım sağlayan bir alıcı önerilip önerilemeyeceği sorgulanmaktadır. Bu amaçla, nicemleme gürültüsü altında çalışan tekli-taşıyıcılı MIMO yapıları için çarpan çizge tabanlı ve çift yönlü karar geribildirimi yöntemini kullanan, düşük karmaşıklığa sahip Ungerboeck tipindeki bir alıcı, söz konusu yapılar için LMMSE tabanlı bir kanal kestirim algoritması ile birlikte önerilmektedir. Önerilen alıcının literatürde benzer karmaşıklığa sahip bir alıcıdan SER ve kullanıcı başına erişilebilir oran kriterleri açısından daha iyi başarım sağladığı gösterilmektedir.

Anahtar Kelimeler: kitlesel MIMO, analogdan-sayısala-dönüştürücü (ADC), nicemleme, bir-bit, aşırı örnekleme, başarım analizi, kanal kestirimi, düşük-çözünürlük, çıkış-yolu, düz-sönümleme, geniş-bant, frekans-seçici sönümleme, tekli-taşıyıcı, dik-frekans-bölümlemeli çoğullama, Ungerboeck, Bussgang dönüşümü, yinelemeli alıcı, indirgenmiş durum, karar geribesleme, sıfır-zorlayıcı, en yüksek oran birleştirme, sıralı LMMSE.

To my parents, spouse and dearest child

ACKNOWLEDGMENTS

First of all, I would like to express my sincere regards and gratitude to my PhD advisor, Prof. Ali Özgür Yılmaz, for his world-class supervision on both technical and non-technical aspects during my whole life as a graduate student. I do not remember any conversation with an offending nature between me and him, especially during the very though periods in my PhD life, which I think should be very rare for such a long period. This certainly reflects his utmost kindness, understanding and confidence that I will succeed, which was a key motivating factor for my PhD journey.

I am also grateful to Prof. Erik G. Larsson from Linköping University, Sweden, for accepting me as a visiting PhD student in his top-tier research group in my PhD thesis subject. I am also thankful to him along with Prof. Emil Björnson and Prof. Håkan Johansson in the same research group for their expert supervision and warm support during my visiting period in Sweden. I consider myself very lucky to work with them as my first international experience in my graduate research life, which contributed to my self-confidence to work in an international environment a lot. I also would like to thank Ziya Gülgün, Özlem Tuğfe Demir and Daniel Verenzuela in the same group for their enjoyable friendship and support during my visit to Sweden and throughout my PhD period.

As one of the most important figures in my graduate life, Dr. Gökhan Muzaffer Güvensen should be mentioned. It has been very intriguing how fast our relationship with him evolved from a student-teaching assistant relationship to a solid friendship that will persist for years since the very early stages of my career as a teaching assistant at METU. It was also interesting to work with him in a technical study that constituted some of the important content of this thesis.

I should also give special credit to my mother Neslihan and father Murat for their continuous support throughout my whole life. I am thrilled to see that they are still so much willing to support me in various ways despite the fact that I am in my 30s. I will

consider myself successful if I can provide at least the same degree of support to my children as I have received from my mother and father. I think that their parenthood has been exemplary to an unquestionable degree.

I would also like to thank Utku Çelebi, Ömer Çayır, Murat Babek Salman, Anıl Kurt, and Esen Özbay for their enjoyable chats and support for my teaching assistant tasks. Utku deserves special credit as he helped me a lot constructing the laboratory for the EE-435 and EE-436 courses.

I should also mention my gratitude to ASELSAN and TUBİTAK for the scholarship they provided throughout my PhD. I would like to thank Dr. Oğuzhan Atak for the guidance he provided for my PhD work as an expert from industry.

Lastly, but not least, I would like to express my gratitude to my wife İpek for all her support during my PhD and for giving birth to our child Uzay.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ACI	Adjacent Channel Interference
ADC	Analog-to-Digital-Converter
AIR	Achievable Rate
APP	A Posteriori Probability
BER	Bit-Error-Rate
BLMMSE	Bussgang Linear Minimum-Mean-Square Error
bpcu	Bits per channel use
BPSK	Binary Phase Shift-Keying
CDF	Cumulative Distribution Function
CFO	Carrier Frequency Offset
CLT	Central Limit Theorem
CMF	Cumulative Mass Function
CP	Cyclic-Prefix
CPRI	Common Public Radio Interface
C-RAN	Cloud Radio Access Network
CSCG	Circularly Symmetric Complex Gaussian
CSI	Channel State Information
DFT	Discrete Fourier Transform
DTFT	Discrete-Time Fourier Transform
EM	Expectation-Maximization
FTSR	Faster-Than-Symbol Rate
GAMP	Generalized Approximate Message Passing
IDFT	Inverse Discrete Fourier Transform
IF	Intermediate Frequency

ISI	Inter-Symbol Interference
i.i.d	Independent identically distributed
LMMSE	Linear Minimum-Mean-Square Error
LPF	Low-Pass Filter
LTE	Long-Term Evolution
MAP	Maximum A Posteriori
MIMO	Multiple-Input-Multiple-Output
MISO	Multiple-Input-Single-Output
ML	Maximum-Likelihood
MLSE	Maximum-Likelihood Sequence Estimation
MMSE	Minimum-Mean-Square Error
MRC	Maximal Ratio Combining
MUI	Multi-User Interference
nMSE	Normalized Minimum-Mean-Square Error
OFDM	Orthogonal-Frequency-Division-Multiplexing
PAPR	Peak-to-Average Power Ratio
PSD	Power Spectral Density
PSK	Phase Shift-Keying
QA-UMPA-BDF	Quantization Aware Ungerboeck Type Message Passing Algorithm with Bidirectional Decision Feedback
QPSK	Quadrate Phase-Shift Keying
RF	Radio Frequency
RRC	Root-Raised-Cosine
SAR	Successive Approximation
SC	Single-Carrier
SER	Symbol-Error-Rate
SIMO	Single-Input Single-Output

SINDR	Signal-to-Interference-Noise-and-Distortion Ratio
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
SPA	Sum-Product Algorithm
U-RSSE	Ungerboeck Reduced-State Sequency Estimation
ZF	Zero-Forcing

CHAPTER 1

INTRODUCTION

Communication has always been very critical for mankind since the very early stages of history. In ancient Greeks and even earlier civilizations, messages sent by human couriers, dogs or pigeons, signaling certain events by fire or smoke, horns or drums are well known by the historians [1]. Nevertheless, there were significant limitations associated with means of communication before the nineteenth century. For example, the speed of communication was very slow, limited by the speed of human couriers, dogs or pigeons. Although pigeons have higher speed for the delivery of the messages compared to other methods, they provided one-way postal service only towards their home from wherever they are. However, communication using electrical signals with the invention of telegraph in the nineteenth century has changed the game in the human communication history. Messages started to be delivered in seconds over thousands of kilometers. It was followed by the invention of telephone, radio and television, fax, internet, which made much faster and reliable communication possible. The invention of mobile phones and popular use of wireless communication have also been a major milestone in the communication history. It has changed the way humanity communicate with each other thanks to the flexibility provided by the wireless technology. However, the demand for faster and reliable communication does not diminish due to the unprecedented amount of information created and transferred each second. In fact, regardless of which period of history is concerned, the following facts will not change in the field of wireless communications [2]:

- There will always be an increased demand for mobile or fixed wireless throughput.

- The amount of the entire electromagnetic spectrum will never increase. And the most desirable frequency bands that has favorable properties regarding propagation through buildings and obstacles in the communication environment constitute only a small portion of this spectrum.
- Therefore, there will always be pressure on communication engineers to make inventions that provide higher spectral efficiency.

For communication systems to accommodate both fast and reliable communication between two communicating nodes with high-spectral efficiency, various techniques are proposed. Even if the most complex algorithms are employed to enable the fastest data rates between two communicating nodes, the performance of the communication systems are limited by environmental and thermal noise and the distortion caused by communication channel between the communicating terminals, which is referred to as fading in the communication literature. To overcome the bottleneck caused by fading, single-input multiple-output (SIMO) systems, in which multiple antennas are employed at the receiver side are proposed to combat fading. An alternative scheme that provides diversity as a means to mitigate fading is multiple-input single-output (MISO) systems, in which space-time block coding techniques are used with multiple antennas are used at the transmitter side. Even if better communication quality is attained by employing multiple antennas either at the transmitter or receiver side to combat fading, it can be shown using Shannon capacity theorem that the capacity of systems having multiple antennas at the transmitter and receiver is limited by the minimum of the number of antennas at the transmitter or receiver side [3]. To obtain much higher data rates between a single transmitter and receiver pair through spatial multiplexing, deploying multiple antennas at the receiver side in addition to the transmitter side, has become very popular in the late 1990s. This scheme is referred to as point-to-point multiple-input multiple-output (MIMO) in the related literature. With point-to-point MIMO, independent data streams can be transmitted at the same time period and the same frequency band. If the MIMO channel is in a rich scattering environment, the number of independent data streams that the channel can accommodate is increased, which results in higher capacity gains over the SIMO, MISO or single-input single-output (SISO) schemes. However, there are certain limitations associated with point-to-point MIMO. Firstly, it requires multiple antennas at the user terminal,

as well as complex radio-frequency (RF) chains per antenna. Therefore, since the capacity of the MIMO channel increases with the minimum of the number of antennas at the user terminal and the base station, the capacity of point-to-point MIMO is not scalable, as it may not be feasible to employ very large arrays with complex RF chains at the user side. Secondly, line-of-sight conditions are stressing when all antennas other than the antennas at the base station in the MIMO system are all confined to a limited space in a single user terminal [2]. When line-of-sight conditions are dominant, this means that there is no rich scattering environment, thus the capacity of the MIMO channel does not increase with the number of antennas in the MIMO system. However, even in the absence of rich scattering environment, multiple-users can be supported by MIMO systems. Even under line-of-sight cases, independence between the channels of each user can be attained as shown in [2, Chapter 7], resulting in a higher capacity for multi-user MIMO. Moreover, compared to point-to-point MIMO, MU-MIMO has better scalability as increasing the number of users in MU-MIMO is easier compared to increasing the number of antennas (and RF chains) for a single user terminal in point-to-point MIMO.

Much more recently, the scaled version of multi-user MIMO, which is referred to as massive MU-MIMO, in which the number of antennas and users are much higher than originally considered for multi-user MIMO systems is proposed. The large number of antennas facilitates very accurate beam alignment, which in turn, results in a highly spectral and energy efficient communication scheme, owing to the very high multi-user interference suppression capability without using any dedicated time-frequency resource for each user. There are also some critical advantages of massive MIMO compared to conventional MU-MIMO. First of all, due to the limited number of base station antennas in multi-user MIMO, precoding at the base station in downlink transmission is not sufficient to cancel inter-user interference, thus the capacity achieving strategy for downlink MU-MIMO requires accurate knowledge of the channel both by the base station and the user terminals. This requires substantial amount of pilots transmitted in both directions, resulting in a large overhead for channel estimation phase. However, in massive MIMO, as precoding in downlink is sufficient to cancel the multi-user interference, only base station needs to know the channel, not the user terminals. This reduces the need for two-way pilot transmission, decreasing the pilot

overhead for channel estimation substantially. Another advantage of massive MIMO compared to MU-MIMO is the simplicity of signal processing. In MU-MIMO, highly complex signal processing is required both at the base station and the user equipment side to obtain an optimal performance. Conversely, close to optimal performance is possible with very simple linear signal processing at the base station side in massive MIMO thanks to the channel hardening that occurs when the number of antennas are large. At the user side, there may be no need for an interference canceling operation as precoding at the base station will be enough to cancel inter-user interference in most cases. For all mentioned reasons, massive MIMO seems to be the most promising scheme to be employed in future communication systems.

Despite the advantages of using a large number of antennas, massive MIMO may also bring about the limitation to use low cost, simple and power efficient hardware per antenna. In that respect, use of low-complexity and low-cost analog-to-digital converters (ADCs) employing a small number of bits has recently been promoted in massive MIMO systems [4], [5]. Another reason for the use of low precision ADCs is to limit the overall power consumption of the massive MIMO communication system employing many ADCs [6]. It was stated in [7] that the addition of each ADC resolution bit increased the power consumption of ADC by 2 to 4 times considering all ADC types. Moreover, the power related Walden's figure of merit for ADCs [8, 9] was also shown to increase substantially for sampling rates higher than 100 MHz in [10]. Since high sampling frequencies will be the case for millimeter wave (mmWave) communication scenarios [11], it is reasonable to use low resolution ADCs to limit the power consumption due to ADCs. Another motivation to use low resolution ADCs is to limit the data rates that will far exceed the rates supported by common public radio interface (CPRI) or enhanced CPRI (eCPRI), especially in cloud radio access network (C-RAN) applications [12]. More specifically, one bit ADCs also have the advantage of not requiring automatic gain control units and having very low hardware complexity [13, 14], which makes it a more appropriate choice for use in massive MIMO systems [5].

Even if low-resolution ADCs have the cost and power efficiency advantage, the trade-off is the increased distortion in the received signal due to quantization noise. Therefore, the design of detectors that minimize the effect of quantization distortion on

massive MIMO communication systems becomes an important task.

In this thesis, the main objective is to answer the question whether the large number of antennas in massive MIMO, which we also refer to as *massive sampling in space* in this thesis, is enough to provide an adequate performance in terms of a reliable communication standpoint in a quantized massive MIMO system. An extension of this question is whether time-domain oversampling in addition to massive sampling in space is able to provide significant gains.

In Chapter 3, we propose a detection scheme that performs oversampling in time and show that significant signal-to-noise ratio (SNR) advantages can be obtained with temporal oversampling even when such an oversampling is made in addition to the massive sampling in space which is inherent to the massive MIMO structure. We see that temporal oversampling provides significant benefits in terms of the error-rate and achievable rate performance of such systems. We also make the performance analysis by deriving analytical bounds on the symbol error-rate (SER) and achievable rate performance of oversampled uplink massive MIMO structures with 1-bit quantization. **The work in Chapter 3 is the first to propose temporal oversampling for quantized massive MIMO systems and to make its performance analysis.**

The proposed detector in Chapter 3 is designed for frequency-flat channels. As practical channels are mostly frequency-selective, the extension of the work in Chapter 3 to frequency-selective channels is important and presented in Chapter 4. We show that temporal oversampling is even more beneficial for frequency-selective channels.

Although the advantages observed when temporal oversampling is performed are remarkable, the zero-forcing (ZF) type detectors proposed in Chapter 3-4 are impractical, whose complexity grows with the cube of block length. In fact, the main purpose of the work in Chapter 3-4 is to show how much performance gains are possible when oversampling is employed, that is, proposing a benchmark detector, rather than a feasible detector. However, in Chapter 5, we propose a low-complexity sequential linear minimum-mean-square error (LMMSE) based detector that employs oversampling for one-bit quantized massive MIMO systems. The complexity of the detector in Chapter 5 changes linearly with block length, which is much less than the complexity of the ZF detector in Chapter 3-4. Despite having much less complexity, the

performance of the detector in Chapter 5 is very similar to the ZF type detectors in Chapter 3-4. Therefore, we illustrate in Chapter 5 that the advantages observed with temporal oversampling can also be obtained with a much lower complexity detector, making the conclusions associated with the advantages of temporal oversampling valid also for detectors with reasonable complexity.

Despite the fact that aforementioned advantages of temporal oversampling in time are promising for quantized massive MIMO systems, the question whether these advantages will be preserved when there is a source of significant interference from an adjacent band remains to be answered. In fact, in none of the studies in the literature that deals with the quantized massive MIMO systems, the effect of an interferer in an adjacent channel is examined, apart from analyzing whether oversampling in time will be beneficial for such systems or not. However, such interference can be at significant levels due to near/far effect in a communication system in which users in the adjacent frequency band may be much closer to the receiver than the users in the desired band, thus, their signal may not be adequately suppressed by the receivers intending to extract the signals in the desired band. In fact, having the dynamic range to mitigate such interference is a key reason for using high-resolution ADCs in current systems [15]. Since distortion is large with low-resolution ADCs, there is a risk that such systems are practically nonoperational. In Chapter 5, we try to find the answer to this question by analyzing the effects of oversampling in time for heavily quantized and orthogonal-frequency-division multiplexing (OFDM) modulated massive MIMO systems for an adjacent channel interference (ACI) scenario under frequency selective fading and channel estimation errors. We also propose an LMMSE based channel estimation algorithm that takes into account the effect of quantization, oversampling and the interference from the adjacent band. **This study is the first to analyze such systems with oversampling ADCs under ACI and imperfect channel state information (CSI).** As a result of the analysis, we show that increasing the number of antennas or temporal oversampling can be very effective for ACI suppression even under heavy quantization for massive MIMO systems.

The proposed detectors using temporal oversampling in the aforementioned chapters are all linear type detectors. However, the optimal detectors for massive MIMO under quantization need not be linear. In Chapter 6, we propose a non-linear Ungerboeck

type detector based on a factor graph constructed over a bidirectional decision feedback algorithm for quantized massive MIMO systems. With this detector, we try to find whether we can achieve a better performance compared to the existing receivers in the literature, by not resorting to any oversampling in time, relying only on the existence of massive sampling in space for massive MIMO systems, by proposing a near optimal receiver. We observe that we are also able to propose a detector showing a better performance compared to the existing detectors of comparable complexity, even if we do not employ any temporal oversampling.

In summary, the thesis is organized as follows. Chapter 3 provides the details of the proposed detector structure that employs temporal oversampling and the corresponding performance analysis for one-bit quantized massive MIMO structures under flat fading. Chapter 4 describes the proposed detector structure for frequency-selective massive MIMO structures with one-bit ADCs and the corresponding performance analysis of such systems. Chapter 5 presents a lower complexity alternative detector to the detectors in Chapter 3 and Chapter 4, which do not experience any performance degradation despite the complexity reduction. Chapter 6 presents our work making the performance analysis of quantized massive MIMO systems under adjacent channel interference with oversampling under perfect or imperfect CSI. Chapter 7 presents the details of the proposed near optimal Ungerboeck type factor-graph based detector with bidirectional decision feedback designed for quantized wideband massive SC-MIMO. Finally, Chapter 8 provides the concluding remarks.

Each chapter (other than Chapters 1-2, which are introductory chapters) in this thesis covers a material that is either published in a paper or submitted for publication. The publications associated with each chapter can be listed as follows:

- **Chapter 3**: A. B. Üçüncü, A. Ö. Yılmaz, "Oversampling in One-Bit Quantized Massive MIMO Systems and Performance Analysis," IEEE Transactions on Wireless Communications, vol. 17, no. 12, pp. 7952-7964, Dec. 2018 [16].
- **Chapter 4**: A. B. Üçüncü, A. Ö. Yılmaz, "Uplink Performance Analysis of Oversampled Wideband Massive MIMO with One-Bit ADCs," Proceedings of 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, 2018, pp. 1-5 [17].

- **Chapter 5**: A. B. Üçüncü, A. Ö. Yılmaz, "Sequential Linear Detection in One-Bit Quantized Uplink Massive MIMO with Oversampling," Proceedings of 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, 2018, pp. 1-5 [18].

- **Chapter 6**: A. B. Üçüncü, E. Björnson, H. Johansson, A. Ö. Yılmaz, E. G. Larsson, "Performance Analysis of Quantized Uplink Massive MIMO-OFDM With Oversampling Under Adjacent Channel Interference," in IEEE Transactions on Communications, vol. 68, no. 2, pp. 871-886, Feb. 2020 [19]. Conference paper version is published in [20].

- **Chapter 7**: A. B. Üçüncü, G. M. Güvensen, A. Ö. Yılmaz, "A Reduced Complexity Ungerboeck Receiver for Quantized Wideband Massive SC-MIMO," in IEEE Transactions on Communications, vol. 69, no 7, pp. 4921-4936, Jul. 2021 [21].

The commercial rights of the material covered in Chapter 3-5 is under protection with an issued United-States (US) patent with number US10447504B1 [22].

CHAPTER 2

FUNDAMENTALS OF QUANTIZED UPLINK MASSIVE MIMO

In this thesis, a frequency-selective single-cell uplink multi-user massive MIMO system with K single-antenna users and M receive antennas with low-resolution ADCs as in Fig 2.1 is examined.

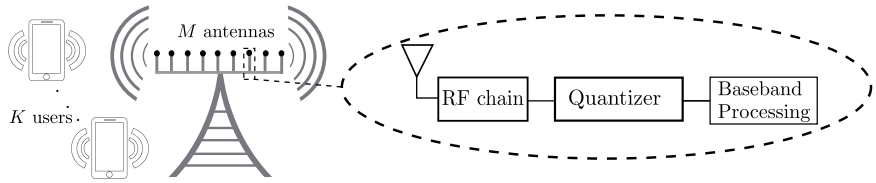


Figure 2.1: Uplink massive MIMO system with low resolution quantizers.

In a massive MIMO system, the number of antennas M is typically much larger than the number of users K. A more detailed picture including the sub-blocks generating the transmitted signals by each user and the sub-blocks associated with each receive antenna in the massive MIMO system for single-carrier modulation is illustrated in Fig. 2.2.

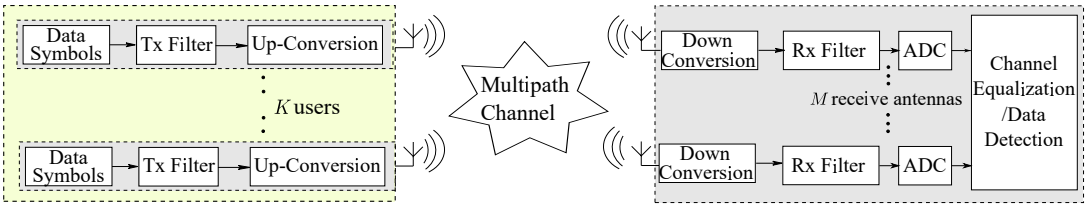


Figure 2.2: Single-carrier multi-user uplink massive MIMO block diagram.

As can be noted in Fig. 2.2, the complex data symbols of each user is fed to the transmit pulse-shaping filter, which is then upconverted to the carrier frequency and transmitted as an analog signal. At the receiver side, the “RX filter" block is assumed to be a pulse matched filter, rather than a channel matched filter, as an analog filter

implementation of channel matched filter requires an adaptive analog filter whose impulse response should be changed when the channel is changed, which is quite complex to implement in practice.

2.1 Notation

The following notation will be used throughout the thesis: b is a scalar, $\mathbf{b}$ is a column vector, b_k is the kth element of $\mathbf{b}$, $\mathbf{B}$ is a matrix, $\mathbf{B}^T$, $\mathbf{B}^*$ and $\mathbf{B}^H$ represents the transpose, conjugate and the Hermitian of a matrix $\mathbf{B}$, respectively. $\mathrm{Re}(.)$ and $\mathrm{Im}(.)$ takes the real and imaginary parts of their operands and $j = \sqrt{-1}$. $\mathbf{0}$ and $\mathbf{I}$ corresponds to a zero column vector and identity matrix with appropriate dimension, respectively. $\mathbf{0}_K$ and $\mathbf{I}_K$ are zero and identity matrices with size $K \times K$. $(\mathbf{C})_{m,n}$ or $[\mathbf{C}]_{(m,n)}$ stands for the element of matrix $\mathbf{C}$ at its mth row and n^{th} column and $|\mathcal{G}|$ represents the cardinality of the set $\mathcal{G}$. $||.||$ represents Euclidean norm, $\mathbf{E}[.]$ is the expectation operator and $Pr(.)$ denotes the probability of the event in its operand. $\mathrm{diag}(\mathbf{C})$ is a diagonal matrix, whose diagonal entries are equal to the diagonal entries of $\mathbf{C}$. Moreover, $\log_2(.)$ is the base-2 logarithm, $\otimes$ is the Kronecker product. Tr[.] is the trace operator. Furthermore, $\mathcal{Q}(.)$ is the Q-function, which is defined as $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-z^2/2} dz$ and $\Phi(.)$ is the standard normal cumulative distribution function (CDF), that is, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz$. $\mathrm{blkToeplitz}(\mathbf{C}, \mathbf{R})$ indicates a block Toeplitz matrix of dimension $M \times K$, whose first row block is matrix $\mathbf{R}$ of size $M_r \times K$ and the first column block is matrix $\mathbf{C}$ of size $M \times K_c$.

2.2 Received Signal Model for Uplink SC-Massive MIMO

The baseband equaivalent transmitted single-carrier signal for the kth user can be expressed as

$$s_k(t) = \sum_{n=1}^{N} x_{n,k} p_c(t - (n-1)T), \tag{2.1}$$

where N is the block length, which is the number of symbols that are processed in a block, $p_c(t)$ is the transmit pulse shaping filter impulse response, $x_{n,k}$ is the complex valued data symbol transmitted by the kth user at the nth symbol interval with unit

average power so that $\mathbf{E}[|x_{k,n}|^2] = 1\ \forall k, n$. T is the symbol period.

The baseband equivalent received signal at the mth antenna can be expressed as

$$r_m(t) = \sum_{k=1}^{K} h'_{m,k}(t) * s_k(t) + w_m(t), \tag{2.2}$$

where $*$ represents the convolution operation and $w_m(t)$ is the filtered additive white complex Gaussian noise process at the mth antenna, whose each sample is a $\mathcal{CN}(0, \sigma_m^2)$ random variable, where $\mathcal{CN}(\mu, \sigma^2)$ represents a circularly symmetric complex Gaussian random variable with mean μ and variance σ^2. Moreover, $h'_{m,k}(t)$ is the continuous-time channel impulse response between the kth user and the mth antenna. In writing (2.2), it is assumed that the bandpass RF filter and the LPF after the mixer, which are considered to be an element of the downconversion sub-block in Fig. 2.2, have a flat response over the frequency bands $[f_c - F_s/2,\ f_c + F_s/2]$ and $[-F_s/2,\ F_s/2]$, f_c being the carrier frequency and F_s being the sampling rate. We will assume integer multiples of symbol duration T for the channel tap delays in this thesis, which is also assumed so in many related studies including one of the pioneer work on quantized massive MIMO [23]. In this case, the received signal in (2.2) can be reexpressed as

$$r_m(t) = \sum_{\ell=0}^{L-1} \sum_{k=1}^{K} \sum_{n=1}^{N} h_{m,k}[\ell] x_{n,k} p_c(t - (n-1)T - \ell T) + w_m(t) \tag{2.3}$$

where L is the number of channel taps, and $h_{m,k}[\ell] = h_{m,k}(\ell T)$, in which $h_{m,u}(t) \triangleq h'_{m,u}(t) * p_c(t)$. The channel taps $h_{m,k}[\ell]$ are generally assumed to be zero-mean circularly symmetric complex Gaussian (CSCG) random variables, corresponding to a Rayleigh fading scenario, and uncorrelated, that is, $\mathbb{E}[h_{m_1,k_1}[\ell_1] h_{m_2,k_2}[\ell_2]^*] = \rho_{k_1}[\ell_1]\delta[\ell_1 - \ell_2]\delta[k_1 - k_2]\delta[m_1 - m_2]$, where $\rho_k[\ell]$ is the power-delay profile of the channel between user k and the receive antennas. This means that we assume a rich scattering environment with wide-sense stationary uncorrelated scattering (WSSUS) channel model for the channels between each user and antenna, which is a result of the assumption that different multipath delays are caused by different scatterers [3]. Moreover, the channels observed by different antennas are also be assumed to be uncorrelated, which is valid when the spacing between the antenna elements are greater than one half wavelength [3] in a dense scattering transmission medium. Furthermore, in dense isotropic scattering environments, when the users terminals are sepa-

rated with a distance greater than one wavelength, it is valid to assume uncorrelated channels observed for each user [2, Section 7.2.1], [24]. Such assumptions for the channel coefficients are also commonly adopted in many studies [23, 25, 26] analyzing massive MIMO with low-resolution quantizers, among others. However, we also consider cases of spatially correlated Rayleigh and Rician fading channels in the work in Chapter 7.

The pulse matched filtered signal at the mth antenna, $d_m(t)$, can be expressed as

$$d_m(t) = \sum_{\ell=0}^{L-1} \sum_{k=1}^{K} \sum_{n=1}^{N} h_{m,k}[\ell] x_{n,k} p(t - (n-1)T - \ell T) + z_m(t), \tag{2.4}$$

where $p(t) = p_c(t) * p_c(-t)$ and $z_m(t) = w_m(t) * p_c(-t)$, $*$ denoting the convolution operation. SR sampling of $d_m(t)$ results in a discrete-time signal model consistent with that in [23] when $p(t)$ is a Nyquist pulse. For the single-carrier scenarios investigated in this thesis, oversampling schemes in Chapter 3-5, the sampling period for the received signal $d_m(t)$ is shorter than the symbol period T.

2.3 Received Signal and System Model for Uplink Massive MIMO-OFDM

For OFDM, the system model for SC-MIMO in Fig. 2.2 is modified as in Fig. 2.3.

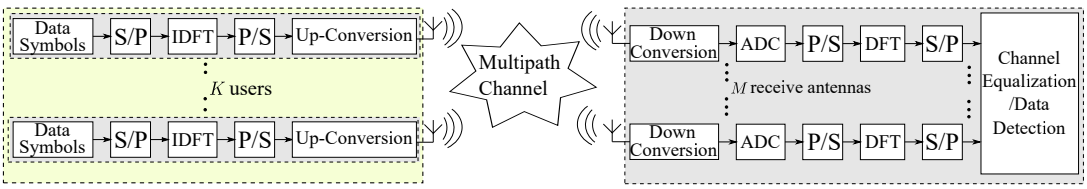


Figure 2.3: Single-carrier multi-user uplink massive MIMO block diagram.

Regarding the received signal model for SC-MIMO in Section 2.2, one modification is that the transmitted symbols $x_{n,k}$ in (2.1), (2.3), (2.4) are replaced with the inverse discrete Fourier transform (IDFT) of the transmitted symbols, namely, $\tilde{x}_{n,k}$, which is found as

$$\tilde{x}_{n,k} = \frac{\rho}{\sqrt{N}} \sum_{u=1}^{N} x_{u,k} e^{j2\pi(n-1)(u-1)/N}, \tag{2.5}$$

where $n = 1, \ldots, N$, ρ is the average transmitted power and $x_{u,k}$ correspond to the transmitted symbol of user k at the uth subcarrier. For simple equalization at the

receiver side, a CP of length L_{cp} is added to the beginning of the OFDM symbol such that $\tilde{x}_{n,k} = \tilde{x}_{N+n,k}$ for $n = -L_{cp} + 1, \ldots, 0$. It is required that $L_{cp} \geq L - 1$, L being the number of channel taps[1]. Therefore, the time index n in (2.1) will be from $-L + 2$ to N. The other modification in (2.1)-(2.4) for OFDM case is that the symbol period T in single-carrier case is replaced by the sampling period T_s in OFDM case. The difference of the oversampling scheme investigated for OFDM case from the single-carrier case in Chapters 3-5 is that the bandwidth of the transmit pulse shaping filter is also increased for OFDM, while oversampling is performed only at the receiver side for single-carrier case in Chapter 3-5. It may seem that the total the total transmission bandwidth for OFDM is also increased with oversampling due to the transmit pulse-shaping filter bandwidth expansion, creating a trade-off between the possible advantages of oversampling and transmission bandwidth. However, it will be explained in Chapter 6 that the transmission bandwidth can be kept the same with the proposed oversampling scheme also for the OFDM case.

2.3.1 Quantization

The process of converting a continous-time signal, which can take infinitely many values, to a discrete signal, which can take a finite number of values, is referred to as *quantization*, and the device that performs this process is called as *quantizer*. To begin with, define the set of quantizer output values as $\mathcal{L} = \{\ell_0, \ell_1, \ldots, \ell_{L'-1}\}$, where $L' = 2^q$ is the number of possible quantizer output values, q being the number of bits of the quantizer. Moreover, the quantization thresholds can also be characterized by the set $\mathcal{B} = \{b_0, b_1, \ldots, b_{L'}\}$, where $-\infty = b_0 < b_1 < \cdots < b_{L'} = \infty$. The quantization of $x \in \mathbb{C}$ is characterized by a function $\mathcal{Q}(.)$ defined as

$$\mathcal{Q}(x) = \ell_{f'(\Re(x))} + j\ell_{f'(\Im(x))}, \tag{2.6}$$

where $f'(\Re(x)) = k \in \{0, 1, \ldots, L' - 1\}$ satisfying $b_k \leq \Re(x) < b_{k+1}$, which is defined similarly for the imaginary part of the quantizer input. If the spacing between the consequtive elements in the quantizer output set $\mathcal{L}$ is the same, the quantizer is a *uniform* quantizer, and the spacing between the output values are referred to as the *step size* Δ of the quantizer. Otherwise, the quantizer is a *non-uniform* quantizer.

[1] L is increased when the sampling rate $(1/T_s)$ is increased as the delay spread of the channels does not change with the sampling rate.

As an example to the uniform quantizers, the possible quantizer output values $\ell_i = \Delta(i - L'/2 + 1/2)$, $i = 0, 1, \ldots, L' - 1$, whereas the quantization thresholds $b_i = \Delta\left(i - \frac{L'}{2}\right)$, $i = 1, 2, \ldots, L' - 1$ for a uniform midrise quantizer ($b_0 = -\infty, b_{L'} = \infty$ as previously specified).

For the extreme case of 1-bit quantizer,

$$f'(x) = \frac{\Delta}{2} sgn(x) + j\frac{\Delta}{2} sgn(x), \tag{2.7}$$

where $sgn(.)$ is the signum function. As a continuous signal at the quantizer output is mapped to a set of finite levels, the signals at the input and output of the quantizer are not equal, resulting in a distortion caused by the quantization process. The distortion caused by the quantization can be less for the non-uniform quantizer, whose quantization thresholds are adjusted according to the statistics of the input signal, than the uniform quantizer. The optimal non-uniform quantizer in terms of the mean-squared quantization error is referred to as a Lloyd-Max quantizer [27, 28]. Uniform quantizers are considered in this thesis, and the extension to the Lloyd-Max quantizer case is left fro future work.

2.3.2 Analog-to-Digital Converters

A basic block diagram for an ADC is presented in Fig. 2.4.

Figure 2.4: Analog-to-digital converter block diagram.

The anti-aliasing filter block filters out the out-of-band (OOB) portion of the received signal. In Chapters 3-5, we assume that the anti-aliasing filters in the ADCs are replaced with pulse-matched filters, which also have a similar function of filtering out the received OOB signals. The anti-aliasing filter block is followed by a sampler block. In Chapters 3-5, the sampling rate of the sampling block is faster than the bandwidth of the anti-aliasing filter block, which is performed to reduce the effects of quantization distortion caused by the following quantizer block of q bits. The details of the quantizer block following the sampling block are presented in Section 2.3.1.

For an ideal ADC, the only source of distortion is the quantization distortion due to the mapping of the continuous input signal to a finite set of discrete levels at the quantizer output. However, in practice there are additional distortion sources due to integral and differential nonlinearity, sampling-time jitter, thermal noise [29]. All these distortions result in an effective signal-to-noise-and-distortion (SNDR) ratio at the quantizer output. From SNDR, effective number of bits (ENOB) of the ADC can be calculated according to the following formula [29]:

$$\text{ENOB} = \frac{\text{SNDR[dB]} - 1.76}{6.02} \tag{2.8}$$

For the case of an ideal ADC and with a sinusoidal input having an amplitude equal to the clipping level of the quantizer, the SNDR can be calculated as follows [29]:

$$\text{SNDR[dB]} = 6.02q + 1.76. \tag{2.9}$$

Therefore, for an ideal quantizer, plugging (2.9) into (2.8), the effective number of bits can be found to be equal the number of quantizer bits. However, if we add distortion sources other than the quantization noise for the calculation of SNDR, the effective number of bits will be lower than the number of ADC bits.

Another important property of the ADCs is their increased power consumption with the number of resolution bits q. If we increase the number of bits, we have about 6 dB improvement in SNDR per added bit according to (2.9). However, the trade-off is the increased power consumption. A metric that relates the power consumption, sampling rate and ENOB is the so-called Walden's figure of merit (FOM_W) defined as [29]:

$$\text{FOM}_\text{W} = \frac{P_s}{2^{\text{ENOB}} f_s}, \tag{2.10}$$

where f_s is the sampling rate and P_s is the power dissipation of the ADC. Another widely adopted metric is the so-called Schreier's figure of merit (FOM_S), defined as

$$\text{FOM}_\text{S} \approx (4^{\text{ENOB}})(10^{0.176})\left(\frac{f_s}{2P_s}\right), \tag{2.11}$$

Both metrics, namely FOM_W and FOM_S, imply that the power consumption due to ADC scales linearly with the sampling rate of the ADC. However, while the power dissipation is doubled with each extra bit according to FOM_W, it is quadrapled according to FOM_S. A recent figure comparing the two figure of merit measures to the power dissipation per sampling rate of the real world ADCs presented in International Solid-State Conference (ISSC) and very-large-scale integration (VLSI) circuit symposium from 1997 to 2019 are provided in Fig. 2.5 [10].

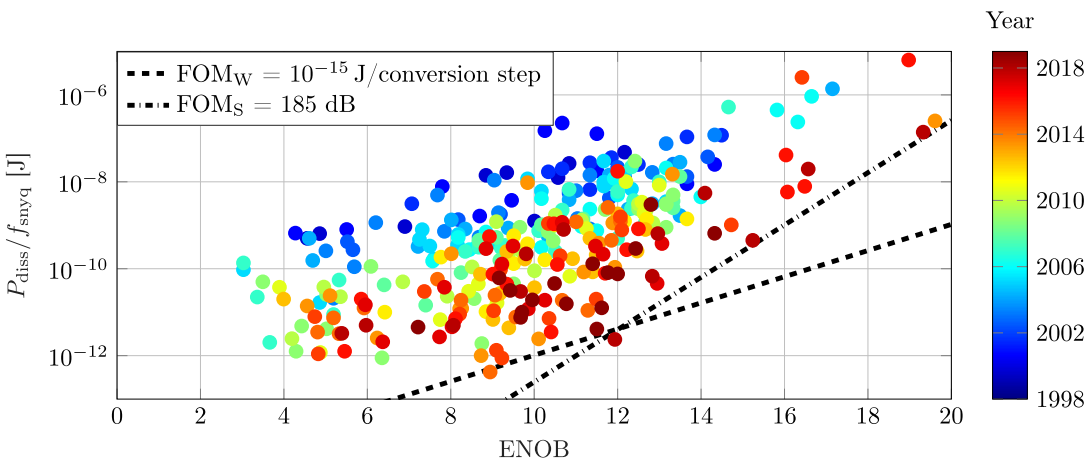


Figure 2.5: Energy (P_s/f_s) versus ENOB for ADCs presented in ISSC and VLSI circuit symposium from 1997 to 2019 [29, Fig. 3.7].

In Fig. 2.5, FOM_W is fixed as $10^{-15} J/\text{conversion step}$, where J stands for the unit in Joules. It gives the energy required to convert an analog signal to a digital signal per each extra quantizer output level. The unit for FOM_W is in $J/\text{conversion step}$ as P_s/f_s is in Joules and 2^{ENOB} is the number of levels in the quantizer. The unit for the other metric FOM_S is also $10^{-15} J/\text{conversion step}$ but it is presented in log domain in Fig. 2.5. As can be noted from Fig. 2.5, the power dissipated by the ADCs, which is indicated with P_{diss}, changes proportional to 4^{ENOB} for high resolution ADCs (ADCs with ENOB>11 bits), which is in-line with the Schreier's figure of merit, whereas P_{diss} scales with 2^{ENOB} for low-to-moderate resolution ADCs. Moreover, the power consumption of the ADCs grows linearly with the sampling rate for sampling rates above 100 MHz. Therefore, employment of power efficient low-resolution

ADCs becomes a necessity for communication scenarios with high-bandwidth. In this thesis, we attempt at answering the question whether we should perform temporal oversampling with low-resolution ADCs. The fact that increasing the sampling rate changes the ADC power consumption linearly, as opposed to the quadratic or quadrupled power consumption increase when the number of resolution bits is incremented, supports our preference of increasing the sampling rate of ADCs in this thesis instead of increasing the ADC bit resolution.

2.3.3 The Bussgang Decomposition

It is possible to express the input-output relation of a quantizer, in fact of any non-linear system, with a statistically equivalent linear relation by employing the Bussgang decomposition or theorem [30,31]. According to the Bussgang theorem and the properties of LMMSE estimation, for a pair of zero-mean jointly complex Gaussian random variables, namely z_m and z_n, with variances σ_m^2 and σ_n^2, and for $r_m = g(z_m)$, where $g(.)$ is any non-linear function, the following relation holds [32], [29]:

$$\mathbb{E}_{r_m,z_n}[r_m z_n^*] = b_m \mathbb{E}_{z_m,z_n}[z_m, z_n^*]. \tag{2.12}$$

Here, $b_m = \mathbb{E}_{z_m}[g(z_m)z_m^*]/\sigma_m^2$. For a zero-mean jointly complex Gaussian random vector $\mathbf{z} \triangleq [z_1 \; z_2 \; \ldots \; z_N]$ with covariance matrix $\mathbf{C}_z$, and $\mathbf{r} \triangleq [r_1 \; r_2 \; \ldots \; r_N]$, (2.12) implies that

$$\mathbf{C}_{\mathbf{rz}} \triangleq \mathbb{E}[\mathbf{rz}] = \mathbf{B}\mathbf{C}_{\mathbf{z}}, \tag{2.13}$$

in which, $\mathbf{B}$ is a diagonal matrix, with diagonal elements being $b_1, b_2, \cdots, b_N$. According to the Bussgang theorem, the output of the non-linear quantizer can be decomposed into two elements, one being a linear function of the input and the other being a distortion that is uncorrelated with the quantizer output, as follows [33]:

$$\mathbf{r} = \mathbf{B}\mathbf{z} + \mathbf{e}. \tag{2.14}$$

It can be shown that if (2.13) holds, the decomposition in (2.14) implies that the distortion term $\mathbf{e}$ is uncorrelated with the quantizer output $\mathbf{r}$. Therefore, the special selection of matrix $\mathbf{B}$ in (2.13) according to the Bussgang theorem results in an uncorrelated

distortion term $\mathbf{e}$ with the quantizer output, which according to the fundamentals of LMMSE estimation implies that the mean squared distortion is minimized. Matrix $\mathbf{B}$ in (2.13) can be calculated for a wide-range of non-linearities, including the case of quantizers. Owing to its capability to represent input-output relations of non-linear systems with a statistically equivalent linear relation with elements having closed-form expressions, the Bussgang decomposition has a wide range of applications for the analysis of the problems containing non-linear elements [29].

CHAPTER 3

OVERSAMPLING IN ONE-BIT QUANTIZED MASSIVE SC-MIMO SYSTEMS AND PERFORMANCE ANALYSIS

In this chapter, we present the performance analysis of one-bit quantized massive single-carrier MIMO (SC-MIMO) systems with oversampling applied in time-domain. Beginning with the construction of a signal model for one-bit quantized massive SC-MIMO systems under flat fading, we make a performance analysis in terms of SER and acheivable rate per user metrics. We propose an upper bound on SER and a lower bound on achievable rate. Chapter 3 continues with simulation results, which verify the accuracy of our analysis and show that temporal oversampling employed in one-bit quantized massive SC-MIMO systems can provide significant SNR gains. Our results establish a tradeoff between oversampling rate and number of antennas.

3.1 Related Works

Owing to their low-cost and power efficiency, 1-bit ADCs have been investigated for various communication schemes. For instance, [6, 14] examined their use in ultra-wideband systems. Moreover, [34, 35] employed 1-bit ADCs for the communication systems that operate in mmWave band. There have also been many studies regarding the use of 1-bit ADCs in MIMO and massive MIMO systems. Some studies related to our work are the ones that have dealt with achievable rate, capacity and error rate performance for 1-bit quantized MIMO structures. For example, [36] found closed form expressions for the achievable rate and SER of 1-bit quantized massive MIMO systems along with results concerning the impact of imperfect CSI on error rate performance. [37] provided achievable rate curves for quantized MIMO systems based

on simulation based results. Moreover, [38] dealt with waveform design optimization in 1-bit quantized massive MIMO systems in presence of spectral constraints to maximize the achievable rate, which was calculated empirically, not analytically. In [39, 40], an analytical expression for the mutual information between the input and output vector of a quantized MIMO system was provided. However, the numerical complexity for the calculation of the analytical expression becomes very high for massive MIMO structures, as the provided analytical expression involves calculation of the probability mass functions (pmf) for every possible set of transmit and quantized receive vector. In [36], since the mutual information and error rate analytical expressions were found between a single transmitted data symbol of one of the users and the estimate of that symbol, these expressions were evaluated in a simpler manner for large number of receive antennas and users. Therefore, we also examine such performance metrics as in [36] for our analysis. Throughout the rest of the chapter, the terms *mutual information* and *achievable rate* are used interchangeably.

In the aforementioned references [36,37,39,40], the analytical expressions were provided for the achievable rate of quantized MIMO systems rather than the capacity of such structures. That is, there was no optimization regarding the input constellation to maximize mutual information. However, in [41], capacity was found for 1-bit quantized MIMO for the low SNR regime, in which the optimal input constellation was shown to be QPSK. Moreover, in [23,30,42–47] quantized MIMO capacity was found using additive quantization noise model (AQNM) with much restraining assumptions of Gaussian input symbols and quantization noise. However, capacity bounds using AQNM with Gaussian quantizer noise was proven to be loose at high SNR in [34]. Moreover, the tight bound at low SNR in AQNM is close to the achievable rate calculated only under the assumption that data symbols have Gaussian distribution, which also constitutes an upper bound on the capacity when the input symbols are from a discrete alphabet. In addition, [34] provided a detailed capacity analysis for a wide range of scenarios with transmitter CSI and 1-bit quantization, in which the capacity for multiple-input single-output (MISO) channel, infinite SNR capacity of single-input multiple-output (SIMO) channel, bounds on infinite and finite SNR capacity of MIMO channel were found and numerical optimization techniques were proposed to design the capacity achieving input distribution for MIMO channel. The downside of

the study in [34] is that the proposed optimization based technique to find the optimal input constellation for 1-bit quantized MIMO systems has to solve 2^{2M} optimization problems, where M is the number of receive antennas, which is infeasible to solve for massive MIMO systems with a large number of receive antennas. An alternative technique to reduce the complexity of solving 2^{2M} problems was also proposed in [34], but the amount of complexity reduction it provides was not discussed. The largest MIMO system examined in the same study is 8×8. Therefore, we limit our focus to examine the achievable rate of our proposed scheme for 1-bit quantized massive MIMO systems without optimizing the input constellation.

In the aforementioned studies, the focus was on the achievable rate, capacity, and error-rate performances of the quantized MIMO structures. Regarding the detection algorithms in quantized MIMO systems, a modified minimum mean-square-error (MMSE) receiver was proposed in [44], which was later extended by [48] through employing a decision feedback equalizer based method for quantized MIMO detection. Another study [49] formulated the nonlinear MMSE receiver operation as a high-dimensional complex optimization problem and presented low-complexity numerical methods to simplify the problem while not providing any performance analysis. Moreover, in [50], maximum-likelihood (ML) detector for 1-bit quantized MIMO was derived and a suboptimal but low complexity ZF detector was shown to maximize an upper bound on the likelihood function of the received observation vector for such systems. In the same study, ZF receiver in quantized MIMO structures was also shown to exhibit comparable error rate performance to the high complexity ML receiver when the number of receive antennas was over 100, which is a possible case for massive MIMO scenarios. Owing to its low complexity compared to ML receivers and competent error rate performance, we also employ a ZF receiver in this chapter.

3.2 Motivations to employ temporal oversampling

By temporal oversampling in time, we refer to a faster than symbol rate (FTSR) sampling at the receiver side. In FTSR sampling, in addition to the samples that are taken at the symbol rate, the received signal is also sampled between the regular sampling points sampled at symbol rate. Note that this scheme is an oversampling

technique applied at the receiver side and should not be confused with the “faster than Nyquist signalling” technique for which the data bearing pulses are sent faster than the Nyquist rate at the transmitter side, which results in intersymbol interference (ISI) between data symbols in time even if the channel is not frequency selective. Since FTSR sampling creates additional signal space dimension, it can be expected to yield better estimates for the transmit signal vector in quantized MIMO structures resulting in better error rate performance. Another intuitive reason to expect a benefit regarding the performance of 1-bit quantized massive MIMO by employing FTSR sampling is due to the phenomenon named "stochastic resonance" where interfering signals enhance recovery of a signal after quantization. By taking additional samples between the regular sampling points, which include ISI, we expect to recover the received signal better under quantization.

3.3 Contributions

With the motivations to employ oversampling at the receiver side, for uplink massive MIMO case, which are mentioned in Chapter 1-2, the following contributions are presented in this chapter:

- We apply FTSR sampling as a novel technique to obtain better achievable rate and SER performance for 1-bit quantized uplink massive MIMO systems. By constructing the signal model for temporally oversampled massive MIMO and expressing the received signal vector in a compact linear model form in Section 3.4, we derive the ZF receiver for temporally oversampled uplink massive MIMO systems with 1-bit ADCs. The proposed ZF receiver that utilizes temporal oversampling provides up to 5 dB SNR gain in terms of the SER and mutual information performance of such systems, both with channel estimation errors and without. Moreover, we show that we can achieve the same error rate performance at a certain SNR level with FTSR sampling with about 200 antennas instead of using 400 receive antennas with no FTSR sampling. This significantly reduces the required form factor for the massive MIMO array without any error rate performance degradation which can be important in terms of the

hardware implementation of such structures. Moreover, as the number of RF components scale linearly with the number of antennas, the total power consumption will also be decreased. In addition, while ADC power consumption is quadratically related with the number of ADC bits, it is only linearly related with oversampling rate [7]. The tradeoff between the total power consumption, power consumption due to ADCs, number of ADC bits, oversampling rate and total number of antennas is worth to be examined thoroughly in a future study.

- We propose a channel estimation algorithm for FTSR case based on the Bussgang linear minimum mean squared error channel estimate proposed in [30] for 1-bit quantized uplink massive MIMO systems. Moreover, we also consider the effect of timing error by assuming that the signals from each user arrive at the receiver side at different time instants.

- We obtain an analytical lower bound on the mutual information between a transmitted symbol and its estimate and an upper bound on SER in 1-bit quantized uplink massive MIMO structures with FTSR sampling operating with a linear receiver for the whole SNR range. We compare the simulation based SER and achievable rate curves with our analytical bounds when ZF and maximal ratio combining (MRC) receivers are used for 1-bit quantized uplink massive MIMO systems with FTSR sampling. We also prove that our bounds on SER and achievable rate are tight at low SNR. Moreover, we observe that the proposed bounds are very close to the results obtained from simulations.

- The bounds that we provide also apply for symbol rate (SR) sampling case, which was investigated in [36]. We show that our bounds are in better consistency with the simulation results than the approximate analytical expressions provided in [36] for the SER and achievable rate for 1-bit quantized uplink massive MIMO structures with ZF type receiver.

- The bounds obtained in this chapter can be found analytically by using arcsine law without resorting to Monte-Carlo techniques to find the conditional covariance matrix of the quantized receive signal vector as performed in the recent study [12] for SR sampling case.

In [51,52], analytical expressions for achievable rate were calculated for oversampled

1-bit quantized SISO case. However, extending them to massive MIMO case results in expressions that will be very complex to calculate, since it will involve calculation of joint pmfs for every possible set of input vectors (the number of pmfs to calculate will be on the order of P^{L+K}, where K is the number of users, P is the order of constellation, L is the length of the pulse shape in discrete samples). Therefore, we provide a different analysis for the achievable rate which only requires the calculation of a single joint pmf of a transmitted data symbol and its estimate.

3.4 Signal Model and CSI Acquisition

We start from the received and pulse-matched filtered signal at the mth antenna in an SC-MIMO system expressed in (2.4) in Chapter 2. This signal can be written for the case of flat fading channel as

$$d_m(t) = \sum_{k=1}^{K} \sum_{n=1}^{N} c_{m,k} x_{n,k} p(t-(n-1)T) + z_m(t), \tag{3.1}$$

where $c_{m,k} = h_{m,k}[0]$ is the channel coefficient between the mth antenna and the kth user, $x_{n,k}$ is the transmitted data symbol of user k at the nth symbol period, $z_m(t)$ and $p(t)$ is the matched filtered pulse shape and noise, as defined in Chapter 2. For ease of demonstration, we define a vector $\mathbf{y}$ as in

$$\mathbf{y} = \begin{bmatrix} [\mathbf{y}^{\mathbf{SR}}]^T & [\mathbf{y}^{\mathbf{OS},1}]^T & [\mathbf{y}^{\mathbf{OS},2}]^T & \cdots & [\mathbf{y}^{\mathbf{OS},\beta-1}]^T \end{bmatrix}^T_{1\times\beta MN}, \tag{3.2}$$

where

$$\mathbf{y}^{\mathbf{SR}} = \begin{bmatrix} y^{SR}_{1,1} & y^{SR}_{1,2} & \cdots & y^{SR}_{1,M} & y^{SR}_{2,1} & \cdots & y^{SR}_{N,M} \end{bmatrix}^T_{1\times MN}, \tag{3.3}$$

$$\mathbf{y}^{\mathbf{OS},b} = \begin{bmatrix} y^{OS,b}_{1,1} & y^{OS,b}_{1,2} & \cdots & y^{OS,b}_{1,M} & y^{OS,b}_{2,1} & \cdots & y^{OS,b}_{N,M} \end{bmatrix}^T_{1\times MN}, \tag{3.4}$$

$b = 1, ..., \beta - 1$ with positive integer oversampling rate β, which is defined as the ratio of the total number of samples to the samples taken at symbol rate. In (3.3) and (3.4),

$$y^{SR}_{i,m} = d_m((i-1)T), \tag{3.5}$$

which corresponds to the samples taken at the symbol rate and

$$y_{i,m}^{OS,b} = d_m((i-1)T + bT/\beta), \tag{3.6}$$

corresponding to the FTSR samples. Furthermore, we also define vectors $\mathbf{x}$ and $\mathbf{n}$ as

$$\mathbf{x} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,K} & x_{2,1} & \cdots & x_{N,K} \end{bmatrix}^T_{1\times NK}, \tag{3.7}$$

$$\mathbf{n} = \begin{bmatrix} [\mathbf{n}^{\mathbf{SR}}]^T & [\mathbf{n}^{\mathbf{OS},1}]^T & [\mathbf{n}^{\mathbf{OS},2}]^T & \cdots & [\mathbf{n}^{\mathbf{OS},\beta-1}]^T \end{bmatrix}^T_{1\times \beta MN}, \tag{3.8}$$

where

$$\mathbf{n}^{\mathbf{SR}} = \begin{bmatrix} n_{1,1}^{SR} & n_{1,2}^{SR} & \cdots & n_{1,M}^{SR} & n_{2,1}^{SR} & \cdots & n_{N,M}^{SR} \end{bmatrix}^T_{1\times MN}, \tag{3.9}$$

$$\mathbf{n}^{\mathbf{OS},b} = \begin{bmatrix} n_{1,1}^{OS,b} & n_{1,2}^{OS,b} & \cdots & n_{1,M}^{OS,b} & n_{2,1}^{OS,b} & \cdots & n_{N,M}^{OS,b} \end{bmatrix}^T_{1\times MN}, \tag{3.10}$$

$b = 1, ..., \beta - 1$, In (3.9) and (3.10), $n_{i,m}^{SR} = z_m((i-1)T)$ and $n_{i,m}^{OS,b} = z_m((i-1)T + bT/\beta)$. In this case, (3.1) can be written in matrix-vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.11}$$

where

$$\mathbf{H} = \begin{bmatrix} [\mathbf{G}_0]^T & [\mathbf{G}_1]^T & [\mathbf{G}_2]^T & \cdots & [\mathbf{G}_{\beta-1}]^T \end{bmatrix}^T, \tag{3.12}$$

$$\mathbf{G}_b = \begin{bmatrix} \mathbf{C}\mathbf{\Lambda}_0^b & \mathbf{C}\mathbf{\Lambda}_{-1}^b & \mathbf{C}\mathbf{\Lambda}_{-2}^b & \cdots & \mathbf{C}\mathbf{\Lambda}_{-(N-1)}^b \\ \mathbf{C}\mathbf{\Lambda}_1^b & \mathbf{C}\mathbf{\Lambda}_0^b & \mathbf{C}\mathbf{\Lambda}_{-1}^b & \cdots & \mathbf{C}\mathbf{\Lambda}_{-(N-2)}^b \\ \mathbf{C}\mathbf{\Lambda}_2^b & \mathbf{C}\mathbf{\Lambda}_1^b & \mathbf{C}\mathbf{\Lambda}_0^b & \cdots & \mathbf{C}\mathbf{\Lambda}_{-(N-3)}^b \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{C}\mathbf{\Lambda}_{N-1}^b & \mathbf{C}\mathbf{\Lambda}_{N-2}^b & \cdots & \mathbf{C}\mathbf{\Lambda}_1^b & \mathbf{C}\mathbf{\Lambda}_0^b \end{bmatrix}_{MN\times NK}, \tag{3.13}$$

$b = 1, ..., \beta - 1$. In (3.13), $\mathbf{C}$ is the matrix, whose element at its mth row and the kth column is equal to $c_{m,k}$. It corresponds to the channel matrix for symbol rate sampling case. Moreover, $\mathbf{\Lambda}_b^n$ is a diagonal matrix of size $K \times K$ whose k^{th} diagonal element is equal to $p(nT + bT/\beta + e_k)$, where e_k is the timing error term associated with the kth user, which is modeled as a zero mean Gaussian random variable of variance σ_e^2, independent of the timing errors of the other users. The matrix $\mathbf{H}$ in (3.12) can be

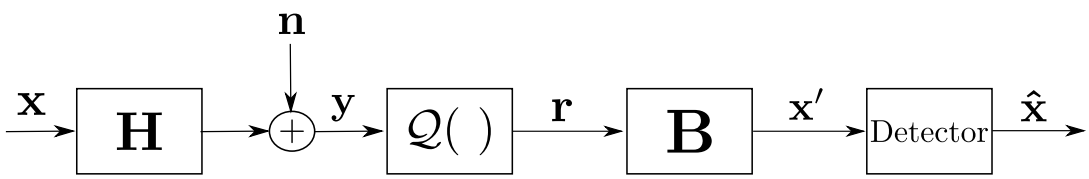


Figure 3.1: Block diagram for overall system

considered to have two parts, namely the part composed of the matrix $\mathbf{G}_0$ and the part that is consisting of $\mathbf{G}_b$, $b = 1, ..., \beta - 1$. The part of $\mathbf{H}$ composed of $\mathbf{G}_0$ corresponds to the relation between the transmitted symbols and the SR samples. It becomes a block diagonal matrix with diagonal elements being the matrix $\mathbf{C}$, if $p(t)$ is selected as a zero-ISI Nyquist pulse. The remaining part of the matrix $\mathbf{H}$ that is composed of $\mathbf{G}_b$'s for $b = 1, ..., \beta - 1$ establishes the relation between the transmitted symbols and the FTSR samples taken at the receiver side. Regarding the effect of the pulse shape, if $p(t)$ decays fast in time, which corresponds to a high roll-off factor case when $p(t)$ is a raised cosine pulse, this will result in the values of the entries of matrix $\mathbf{G}_b$ decreasing fast as they become distant to the main diagonal of matrix $\mathbf{G}_b$.

In the case that 1-bit quantized version of the received signal vector $\mathbf{y}$ is taken into account, the signal model in (3.11) becomes

$$\mathbf{r} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}(\mathbf{H}\mathbf{x} + \mathbf{n}), \tag{3.14}$$

where $\mathcal{Q}(\mathbf{y})$ =sgn$(\mathrm{Re}(\mathbf{y})) + jsgn(\mathrm{Im}(\mathbf{y}))$, sgn(.) being the signum function. Without loss of generality, we assume MRC or ZF type linear receivers. In that case, the soft estimate for the transmitted symbol vector $\mathbf{x}'$ can be found as [36, 53]

$$\mathbf{x}' = \mathbf{B}\mathbf{r}, \tag{3.15}$$

where $\mathbf{B}$ is a linear receive filter. For MRC and ZF type receivers $\mathbf{B} = \hat{\mathbf{H}}^H$ and $\mathbf{B} = (\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^H$, respectively, where $\hat{\mathbf{H}}$ is the estimate for the oversampled channel matrix $\mathbf{H}$. The details of how to obtain $\hat{\mathbf{H}}$ will be discussed in Section 3.4.1. The hard symbol estimate for the transmitted symbol vector $\hat{\mathbf{x}}$ is found by mapping the elements of $\mathbf{x}'$ to the minimum distance constellation point. The overall system model is illustrated with a block diagram in Fig.3.1.

3.4.1 CSI Acquisition

Owing to the fact that the channel coefficients $c_{m,k}$ and the timing errors associated with each user are not known at the receiver side, the optimal way to estimate $\mathbf{H}$ requires the joint estimation of the channel coefficients $c_{m,k}$ and the timing errors e_k. In fact, since there is no linear relationship between the timing errors e_k and the unquantized observations, their estimation, even with symbol rate sampling, is not a straightforward task which is not considered in any of the studies regarding 1-bit quantized massive MIMO. However, if e_k were known, $\mathbf{H}$ could be directly found from $\mathbf{C}$ using (3.12) and (3.13). Therefore, instead of the joint estimation of e_k's and channel coefficients, we propose a suboptimal strategy for the estimation of $\mathbf{H}$ in which e_k's are assumed to be equal to zero although they are not. In such a case, estimation of matrix $\mathbf{H}$ boils down to the estimation of matrix $\mathbf{C}$, whose estimation from the symbol rate samples is a well studied task [30, 36, 53]. Among the methods in [30, 36, 53], we adopt the Bussgang-based linear mean squared error (BLMMSE) channel estimator in [30]. We assume that all users are transmitting pilot sequences of τ symbols to the receiver, which yields [30]

$$\mathbf{Y_{pilot}} = \mathbf{C}\mathbf{\Phi}^{\mathbf{T}} + \mathbf{N_{pilot}} \tag{3.16}$$

where $\mathbf{Y_{pilot}}$ is a matrix of size $M \times \tau$ representing the received signal in the pilot symbol transmission phase, composed of symbol rate samples, whose each row corresponds to the received signal for a certain antenna during the pilot transmission phase, $\mathbf{\Phi}$ is the pilot matrix of size $\tau \times K$, whose each column corresponds to the pilot sent by each user and $\mathbf{N_{pilot}}$ is the noise matrix, whose elements are independent zero mean complex Gaussian random variables with variance σ_n^2. It is important that the pilot sequences transmitted by each user are orthogonal, thus $\mathbf{\Phi}^H\mathbf{\Phi} = \tau\mathbf{I}$. Therefore, DFT vectors can be appropriate selections for the pilot sequences. For such a selection, $\mathbf{\Phi}$ is equal to the first K columns of the DFT matrix of size $\tau \times \tau$. Denoting $Vec(\hat{\mathbf{C}})$ by $\hat{\underline{\mathbf{c}}}$, where $Vec(.)$ is the matrix to vector conversion operator, it can be found as [30]

$$\hat{\underline{\mathbf{c}}} = \tilde{\mathbf{\Phi}}^{\mathbf{H}}\mathbf{F}_{\mathbf{pilot}}^{-\mathbf{1}}\mathbf{r_{pilot}}, \tag{3.17}$$

where $\mathbf{r_{pilot}} = \frac{1}{\sqrt{2}}\mathcal{Q}\left(Vec\left(\mathbf{Y_{pilot}}\right)\right)$,

$$\tilde{\mathbf{\Phi}}^{\mathbf{H}} = \mathbf{A}\bar{\mathbf{\Phi}}, \tag{3.18}$$

in which $\bar{\mathbf{\Phi}} = \mathbf{\Phi} \otimes \mathbf{I}_M$ and

$$\mathbf{A} = \sqrt{\frac{2}{\pi}} \left(\left(\mathbf{\Phi}\mathbf{\Phi}^H \otimes \mathbf{I}_\tau \right) + \sigma_n^2 \mathbf{I}_{ML} \right)^{-0.5}. \tag{3.19}$$

Moreover, $\mathbf{F}_{\mathbf{pilot}}$ in (3.17) is the quantized received signal covariance matrix for the pilot phase, which can be found as [30]

$$\begin{aligned} \mathbf{F}_{\mathbf{pilot}} = \frac{2}{\pi} \Big(& \operatorname{asin} \left(\mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}} \operatorname{Re}(\mathbf{G}_{\underline{\mathbf{y}}^{(\mathbf{p})}}) \mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}} \right) \\ & + j \operatorname{asin} \left(\mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}} \operatorname{Im}(\mathbf{G}_{\underline{\mathbf{y}}^{(\mathbf{p})}}) \mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}} \right) \Big), \end{aligned} \tag{3.20}$$

in which $\mathbf{G}_{\underline{\mathbf{y}}^{(p)}} = \left(\left(\mathbf{\Phi}\mathbf{\Phi}^H \otimes \mathbf{I}_\tau \right) + \sigma_n^2 \mathbf{I}_{ML} \right)$, $\mathbf{K}_{\underline{\mathbf{y}}^{(p)}} = \operatorname{diag}\left(\mathbf{G}_{\underline{\mathbf{y}}^{(p)}} \right)$.

When $\hat{\underline{\mathbf{c}}}$ is found from (3.17), which means that $\hat{\mathbf{C}}$ is obtained, $\hat{\mathbf{H}}$ can be computed using (3.12) and (3.13) by replacing zero values for the timing error terms e_k.

3.5 SER and Achievable Rate Analysis of Oversampled Massive MIMO

In this section, we derive analytical expressions for the SER and mutual information for the FTSR sampled and 1-bit quantized uplink massive MIMO. The SER taking into account the mth element of the transmitted data vector $\mathbf{x}$, namely x_m, can be found as

$$\text{SER} = \mathbf{E}_{\mathbf{H}} \left[\sum_{\hat{x}_m \neq x_m} \sum_{x_m} p(\hat{x}_m | x_m, \mathbf{H}) p(x_m) \right], \tag{3.21}$$

where $\hat{x}_m$ is the hard symbol estimate of the transmitted symbol x_m and $p(x_m)$ is the pmf of x_m. Although SER in (3.21) is calculated for a single element of $\mathbf{x}$, namely x_m, the SER taking into account all elements of the transmitted data vector $\mathbf{x}$ except the ones that are close to the beginning and the end of the transmitted data block can also be found using (3.21) since the channel coefficients for different symbol intervals and users, which are assumed to have the same average transmitted power, have identical distributions. The reason for the symbols at the edges of the transmitted block may have different SER values is owing to the fact that the number of FTSR samples that are supposed to refine the estimates are truncated thus limited for the symbols at the block edges.

Considering the structure of $\mathbf{x}$ defined in (3.7), if x_m is the mth element of $\mathbf{x}$, it belongs to the kth user, where $k = \text{mod}\ (m-1, K) + 1$ and the nth symbol interval, where $n = \lfloor \frac{m-1}{K} \rfloor + 1$. Here $\lfloor\ .\ \rfloor$ is the floor function that gives the largest integer less than its operand and mod (a, p) yields the remainder after division of a by p. Furthermore, the mutual information between the transmitted symbol x_m and its hard estimate $\hat{x}_m$, namely $I(x_m; \hat{x}_m)$, can be calculated as

$$I(x_m; \hat{x}_m) = \mathbf{E_H}\left[\sum_{x_m}\sum_{\hat{x}_m} p(\hat{x}_m|x_m, \mathbf{H})p(x_m)\log_2\frac{p(\hat{x}_m|x_m, \mathbf{H})}{p(x_m)}\right]. \tag{3.22}$$

To be able to employ (3.21) and (3.22) for SER and mutual information calculation, the pmf $p(\hat{x}_m|x_m, \mathbf{H})$ needs to be found. To find $p(\hat{x}_m|x_m, \mathbf{H})$, we should obtain the probability density function (pdf) $f(x'_m|x_m, \mathbf{H})$. For the SR sampling case, it has been shown that $f(x'_m|x_m, \mathbf{H})$ can be approximated by a normal distribution whose mean $\mathbf{E}[\hat{x}_m|x_m, \mathbf{H}]$ and variance $\text{Var}(\hat{x}_m|x_m, \mathbf{H})$ has been found approximately [36]. We will also show that $f(x'_m|x_m, \mathbf{H})$ can also be approximated by a Gaussian distribution for the oversampled case and extend the derivation in [36] to find $\mathbf{E}[\hat{x}_m|x_m, \mathbf{H}]$. To find $\text{Var}(\hat{x}_m|x_m, \mathbf{H})$ we use a completely different approach than that in [36], since the approach in [36] is valid only with SR sampling. The resulting analytical expression that we derive for the general oversampled case can also be used for the SR sampling case, which yields SER and mutual information values with better accuracy than that derived in [36]. The reason is that more terms are taken into account to find the variance of the normal distributed conditional random variable $\hat{x}_m$ conditioned on $x_m, \mathbf{H}$, which will be explained later in this section. Once we find both $\mathbf{E}[\hat{x}_m|x_m, \mathbf{H}]$ and $\text{Var}(\hat{x}_m|x_m, \mathbf{H})$, we can find $p(\hat{x}_m|x_m, \mathbf{H})$ from $f(x'_m|x_m, \mathbf{H})$ using Q-functions. To start with, consider the pth element of the unquantized observation vector $\mathbf{y}$, namely y_p. Using (3.11), it can be expressed as

$$y_p = \sum_{j=1}^{NK} h_{p,j}x_j + n_p, \tag{3.23}$$

where x_j and n_p are the jth and p^{th} element of transmitted data vector $\mathbf{x}$ and $\mathbf{n}$. Moreover, $h_{p,j}$ is the element of $\mathbf{H}$ at p^{th} row and j^{th} column. The mean of y_p conditioned on x_m and $\mathbf{H}$ can be found as

$$\mathbf{E}[y_p|x_m, \mathbf{H}] = h_{p,m}x_m \tag{3.24}$$

since $\mathbf{E}[x_j] = 0$ for $j \neq m$ and $\mathbf{E}[n_p] = 0$. Define $v_p \triangleq \sum_{j \neq m} h_{p,j} x_j$. Since all x_j's are independent, which makes $h_{p,j} x_j$ terms in the v_p expression independent for a given $\mathbf{H}$ and each $h_{p,j} x_j$ term has a finite variance and zero mean for $j \neq m$, the distribution of the random variable v_p conditioned on x_m and $\mathbf{H}$ converges to Gaussian by the Central Limit Theorem (CLT) [36], [54] as K is considered to be large. It will be zero mean and its variance $\sigma^2_{v_p} = \sum_{j \neq m} |h_{p,j}|^2$ since $\mathbf{E}[|x_j|^2] = 1$. Considering the definition of v_p, (3.23) can be rewritten as

$$y_p = h_{p,m} x_m + v_p + n_p. \tag{3.25}$$

Therefore, the pdf of y_p conditioned on x_m and $\mathbf{H}$, namely $f(y_p | x_m, \mathbf{H})$, satisfies

$$f(y_p | x_m, \mathbf{H}) \approx \frac{1}{\pi \sigma^2_{p,m}} exp\left(\frac{-||y_p - \mu_{p,m}||^2}{\sigma^2_{p,m}} \right) \tag{3.26}$$

where $\mu_{p,m} = h_{p,m} x_m$ and $\sigma^2_{p,m} = \sum_{j \neq m} |h_{p,j}|^2 + \sigma^2_n$. In addition, when the SNR is low, that is, when n_p is dominant, the approximation in (3.26) will be valid, regardless of K being large and v_p having normal distribution or not. The 1-bit quantized version of y_p can be denoted as r_p which corresponds to the pth element of the vector $\mathbf{r}$ defined in (3.14). The mean of r_p, which can be represented as μ_p, conditioned on x_m and $\mathbf{H}$ can be found as

$$\mu_p = (1+j)\kappa_{1p} + (-1+j)\kappa_{2p} + (1-j)\kappa_{3p} + (-1-j)\kappa_{4p}, \tag{3.27}$$

$\kappa_{1p} = Pr(r_p = 1 + j | x_m, \mathbf{H})$, $\kappa_{2p} = Pr(r_p = -1 + j | x_m, \mathbf{H})$, $\kappa_{3p} = Pr(r_p = 1 - j | x_m, \mathbf{H})$, $\kappa_{4p} = Pr(r_p = -1 - j | x_m, \mathbf{H})$. Considering (3.26), for $x_m = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} j$, κ_{1p} can be found as

$$\kappa_{1p} = Q\left(-\frac{\sqrt{2}\,\mathrm{Re}(h_{p,m} x_m)}{\sqrt{\sigma^2_n + \sum_{j \neq m} |h_{p,j}|^2}} \right) \times Q\left(-\frac{\sqrt{2}\,\mathrm{Im}(h_{p,m} x_m)}{\sqrt{\sigma^2_n + \sum_{j \neq m} |h_{p,j}|^2}} \right), \tag{3.28}$$

where $Q(.)$ is the Q-function. The values of κ_{2p}, κ_{3p} and κ_{4p} can be found similarly. It will be clear shortly where μ_p values will be used.

At this point, consider the mth element of the soft symbol estimate vector $\mathbf{x}'$, which can be denoted as x'_m. Using (3.15) it can be expressed as

$$x'_m = \mathbf{b}^T_m \mathbf{r}, \tag{3.29}$$

where $\mathbf{b}_m^T$ is a row vector equal to the mth row of the linear receive filter matrix $\mathbf{B}$ in (3.15). (3.29) can be rewritten as

$$x'_m = \sum_{p=1}^{MN\beta} b_{m,p} r_p \tag{3.30}$$

$$= s_m + \sum_{p=1}^{MN\beta} b_{m,p}\mu_p, \tag{3.31}$$

where $s_m \triangleq \sum_{p=1}^{MN\beta} b_{m,p}(r_p - \mu_p)$ and $b_{m,p}$ is the element of receive filter matrix $\mathbf{B}$ on the mth row and p^{th} column. Considering the expression for s_m, one can see that it is composed of a summation of many terms of finite variance. Although there is some correlation among r_p, thus the summation terms, it has been shown in [36] with empirical results that CLT can be applied and s_m is approximately normally distributed with SR sampling. We also make a similar assumption that s_m approximately has normal distribution also for the oversampled case based on our empirical observations so that

$$f(x'_m|x_m, \mathbf{H}) \approx \frac{1}{\pi\sigma^2_{x'_m}} exp\left(\frac{-||x'_m - \mu_{x'_m}||^2}{\sigma^2_{x'_m}}\right), \tag{3.32}$$

where $\mu_{x'_m} = \sum_{p=1}^{MN\beta} b_{m,p}\mu_p$, $\sigma^2_{x'_m}$ is the conditional variance of x'_m and $f(x'_m|x_m, \mathbf{H})$ is the pdf of x'_m conditioned on x_m and $\mathbf{H}$. The assumption that $f(x'_m|x_m, \mathbf{H})$ is a Gaussian pdf is valid when M is large, which is the case for a massive MIMO scenario (in our case, we take $M = 400$). Assuming that $f(x'_m|x_m, \mathbf{H})$ is a Gaussian pdf will result in the analytically calculated $I(x_m, \hat{x}_m)$ values to be lower than their exact value, but this effect will be insignificant as M is large. Since μ_p can be found from (3.27), we can find the mean of x'_m conditioned on x_m and $\mathbf{H}$ whose expression is given in (3.32). What remains is to find $\sigma^2_{x'_m}$. Considering (3.30), $\sigma^2_{x'_m}$ can be calculated by using the variance of sum formula [55] as

$$\sigma^2_{x'_m} = \sum_{p=1}^{MN\beta} |b_{m,p}|^2 \sigma^2_{r_p} + \sum_{(p,p'),p\neq p'} Cov(b_{m,p}r_p, b_{m,p'}r_{p'}), \tag{3.33}$$

where $\sigma^2_{r_p}$ is the variance of the random variable r_p and $Cov(x, y) = \mathbf{E}[xy^*] - \mathbf{E}[x]\mathbf{E}[y^*]$. Note that the correlation among $r_p's$ are explicitly taken into account in (3.33), which was not considered for the approximate analytical expressions provided in [36] for SR sampling case. In fact, there exists a correlation between the observations from different antennas for given $\mathbf{H}$ case, which is proven in the following proposition.

Proposition 1: There exists a correlation between the quantized observations from different antennas when $\mathbf{H}$ *and* x_m *is given, that is* $Cov(r_p, r'_p|x_m, \mathbf{H}) \neq 0 \forall p \neq p$ *.*

Proof:

$$Cov(y_p, y'_p|x_m, \mathbf{H}) = \mathbf{E}[y_p(y'_p)^*|x_m, \mathbf{H}] - \mathbf{E}[y_p|x_m, \mathbf{H}]\mathbf{E}[y'_p|x_m, \mathbf{H}]^* \tag{3.34}$$

$$= \sum_{j=1, j\neq m}^{NK} h_{p,j}h^*_{p',j} + h_{p,m}h^*_{p',m}|x_m|^2 - h_{p,m}h^*_{p',m}|x_m|^2 \tag{3.35}$$

$$= \sum_{j=1, j\neq m}^{NK} h_{p,j}h^*_{p',j} \neq 0, \tag{3.36}$$

which implies that y_p and $y_{p'}$ are also correlated when $x_m, \mathbf{H}$ are given. As y_p and $y_{p'}$ are shown to be correlated, their quantized versions, namely r_p and $r_{p'}$, also become correlated. ■

It will be seen in Section 3.6 that this will create significant discrepancy between the empirical SER or achievable rate results for $\beta = 1$ and the analytical expressions provided in [36]. To find $\sigma^2_{x'_m}$, (3.33) which can be written in a more compact form as follows:

$$\sigma^2_{x'_m} = \mathbf{b}_m^T \mathbf{\Gamma_{rr}} \mathbf{b}_m^*, \tag{3.37}$$

where $\mathbf{\Gamma_{rr}} = \mathbf{E}[(\mathbf{r} - \mathbf{E}[\mathbf{r}])(\mathbf{r} - \mathbf{E}[\mathbf{r}])^H]$ is the covariance matrix for vector $\mathbf{r}$. Therefore, it will suffice to find an analytical expression for $\mathbf{\Gamma_{rr}}$ to calculate $\sigma^2_{x'_m}$. Once it is calculated, along with the calculated mean value of x'_m given x_m and $\mathbf{H}$, the pdf of x'_m conditioned on x_m and $\mathbf{H}$, $f(x'_m|x_m, \mathbf{H})$, can be found from (3.32). Using Q-functions, pmf of the hard symbol estimate $p(\hat{x}_m|x_m, \mathbf{H})$ can easily be found from $f(x'_m|x_m, \mathbf{H})$. When $p(\hat{x}_m|x_m, \mathbf{H})$ is obtained, SER and mutual information $I(x_m, \hat{x}_m)$ can be obtained from (3.21) and (3.22).

To calculate $\mathbf{\Gamma_{rr}}$, we first define the vectors $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{y}}$ as follows:

$$\tilde{\mathbf{r}} = \begin{bmatrix} \mathrm{Re}(\mathbf{r}) \\ \mathrm{Im}(\mathbf{r}) \end{bmatrix}, \tilde{\mathbf{y}} = \begin{bmatrix} \mathrm{Re}(\mathbf{y}) \\ \mathrm{Im}(\mathbf{y}) \end{bmatrix}. \tag{3.38}$$

Owing to the approximation in (3.26), each element of $\mathbf{y}$ is Gaussian distributed, thus $\tilde{\mathbf{y}}$ is also Gaussian, which makes the arcsine law to be applicable [56] in that respect.

Therefore, the following matrix equation can be written [45, 56].

$$\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}} = \frac{2}{\pi}\left[\arcsin\left(\mathrm{diag}(\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}})^{-\frac{1}{2}}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}\mathrm{diag}(\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}})^{-\frac{1}{2}}\right)\right]. \tag{3.39}$$

In (3.39), $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}} = \mathbf{E}[\tilde{\mathbf{r}}\tilde{\mathbf{r}}^H]$ and $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = \mathbf{E}[\tilde{\mathbf{y}}\tilde{\mathbf{y}}^H]$. However, (3.39) is valid only when $\tilde{\mathbf{r}}$ is zero mean. If $\tilde{\mathbf{r}}$ were zero mean we would be able to find $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}$ from (3.39) and then find $\mathbf{\Gamma}_{\mathbf{rr}} = \mathbf{R}_{\mathbf{rr}} = \mathbf{E}[\mathbf{r}\mathbf{r}^H]$ from $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}$. However, assuming that $x_m = 0$, in which case (3.39) can be used since $\tilde{\mathbf{r}}$ will be zero mean, we show in Lemma 1 that the conditional variance that we find for $x_m^{'}$ conditioned on $x_m = 0$, which we refer to as $\mathrm{Var}(x_m^{'}|x_m = 0, \mathbf{H})$, is larger than $\sigma^2_{x_m^{'}}$, thus will constitute an upper bound on $\sigma^2_{x_m^{'}}$.

Lemma 1: $\mathrm{Var}(x_m^{'}|x_m = 0, \mathbf{H}) > \sigma^2_{x_m^{'}}$.

Proof: See Appendix A.1. ■

The details for how to find $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ for the case when $x_m = 0$ to employ (3.39) to obtain $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}$ and $\mathrm{Var}(x_m^{'}|x_m = 0, \mathbf{H})$ are presented in Appendix A.2. We also show in Lemma 2 that when we replace $\sigma^2_{x_m^{'}}$ by $\mathrm{Var}(x_m^{'}|x_m = 0, \mathbf{H})$ in (3.32) to find $f(x_m^{'}|x_m, \mathbf{H})$ and $p(\hat{x}_m|x_m, \mathbf{H})$, then utilize (3.21) to find SER, this will constitute an upper bound on the exact SER value.

Lemma 2: If $\mathrm{Var}(x_m^{'}|x_m = 0, \mathbf{H})$ *is used in place of* $\sigma^2_{x_m^{'}}$ *the SER value calculated using (3.21) will yield an upper bound on the actual SER.*

Proof: See Appendix A.1. ■

Therefore, if we denote the pmf of $\hat{x}$ conditioned on x_m and $\mathbf{H}$ that is found using the upper bound on $\sigma^2_{x_m^{'}}$ by $p_u(\hat{x}_m|x_m, \mathbf{H})$, the SER expression in (3.21) can be upper bounded as

$$\mathrm{SER} \leq \mathbf{E}_{\mathbf{H}}\left[\sum_{\hat{x}_m \neq x_m}\sum_{x_m} p_u(\hat{x}_m|x_m, \mathbf{H})p(x_m)\right]. \tag{3.40}$$

Next, we prove that $I(x_m; \hat{x}_m)$ that we calculate using the upper bound on $\sigma^2_{x_m^{'}}$ will yield a lower bound on $I(x_m; \hat{x}_m)$.

Lemma 3: If $\mathrm{Var}(x_m^{'}|x_m = 0, \mathbf{H})$ *is used in place of* $\sigma^2_{x_m^{'}}$, $I(x_m; \hat{x}_m)$ *calculated using (3.22) will yield a lower bound on the actual* $I(x_m; \hat{x}_m)$.

Proof: See Appendix A.3. ■

Then, $I(x_m;\hat{x}_m)$ in (3.22) can be lower bounded as

$$I(x_m;\hat{x}_m) \geq \mathbf{E_H}\left[\sum_{x_m}\sum_{\hat{x}_m} p_u(\hat{x}_m|x_m,\mathbf{H})p(x_m)\log_2\frac{p_u(\hat{x}_m|x_m,\mathbf{H})}{p(x_m)}\right]. \tag{3.41}$$

We also prove that the bounds that we find are tight at low SNR. We prove this by showing that the variance that we find in our analysis, namely $\mathrm{Var}(x'_m|x_m=0,\mathbf{H})$, will converge to the exact value of $\sigma^2_{x'_m}$ in Lemma 4, thus $p_u(\hat{x}_m|x_m,\mathbf{H})$ will be equal to $p(\hat{x}_m|x_m,\mathbf{H})$. Therefore, the RHS of (3.40) and (3.41) becomes equal to RHS of (3.21) and (3.22), which means that the bounds in (3.40) and (3.41) are tight.

Lemma 4: $\mathrm{Var}(x'_m|x_m=0,\mathbf{H}) \to \sigma^2_{x'_m}$ *when* $SNR \to 0$

Proof: See Appendix A.4. ■

Note that arcsine rule in (3.39) is valid only when the quantizer input is zero mean. However, the quantizer input is not zero mean for the condition when x_m and $\mathbf{H}$ are given. Therefore, Lemma 1-3 in this section are important to resolve this discrepancy. Utilization of the arcsine rule in (3.39) is critical since it is important to obtain the matrix $\mathbf{\Gamma_{rr}}$ to calculate $\sigma^2_{x'_m}$, which is essential to calculate SER and achievable rate values. Although we do not arrive at an exact calculation of SER and achievable rate, with the results in Lemma 1-3, we can utilize the arcsine rule to obtain an upper bound on SER as in (3.40) and a lower bound on achievable rate as in (3.41). Moreover, it is also important that the provided bounds will be close to the simulated values for low SNR as they are proved to be tight in Lemma 4 when SNR is low. Note also that the bounds provided on the error rate and achievable rate in this section are in fact approximate bounds owing to the approximations in (3.26) and (3.32).

3.6 Simulation Results

Number of receive antennas (M) and users (K) are taken to be 400 and 20, respectively. Moreover, the block length (N) is selected to be 10 symbols and the oversampling rate (β) is chosen between 1 and 8. The simulation based SER performance plots for the FTSR sampled massive MIMO with 1-bit quantization is obtained by

taking the average of 100 channel matrix, noise and symbol vector triplets. The simulation based mutual information between x_m and $\hat{x}_m$ is found by approximating the conditional probabilities involved in (3.22) through Monte-Carlo simulations which are also performed over 100 channel matrix, noise and symbol vector triplets. For the analytical SER plots, we find the transition probabilities $p_u(\hat{x}_m|x_m, \mathbf{H})$ in (3.40) relying on (3.32) and find the analytical SER using (3.40) for a certain channel matrix realization and the averaging with respect to channel matrix is performed over 100 channel matrix realizations. Note that the provided analytical SER calculation is in fact an upper bound on the actual SER according to Corollary 2. Once the transition probabilities $p_u(\hat{x}_m|x_m, \mathbf{H})$ are found for a certain channel matrix realization, the mutual information between x_m and $\hat{x}_m$ are found using (3.41), which is also averaged over 100 channel matrix realizations. The symbol transmitted in the 5th symbol interval by the 20th user is considered, which is x_{100}. Therefore, $m = 100$. Since the analytical SER is found based on $p_u(\hat{x}_m|x_m = x_{100}, \mathbf{H})$, $x_m = x_{100}$ being one of the symbols transmitted at the 5th symbol interval, the errors that are made only at the 4th, 5th and the 6th symbol intervals, which are around the center of the transmitted block of length 10, are taken into account for the empirical SER curves. In fact, this provides a better characterization of the SER for longer block lengths in comparison to the case that the errors for the symbols at the block edges are taken into account. To ensure a reasonable complexity for the calculation of the upper bound on $\sigma^2_{x'_m}$, the correlation matrices in (3.39) are constructed by only considering the correlation among the SR samples corresponding to the 5th symbol interval and the neighboring FTSR samples that are taken half symbol duration before and after the SR sampling point of the 5th symbol interval since the employed root-raised-cosine (RRC) pulse is assumed to decay to insignificant levels after 2 symbol durations. The SNR values on the plots correspond to per-antenna SNR or $1/\sigma^2_m$, which is taken the same for all receive antennas.

The SER performance and achievable rate per user for the MRC receiver is omitted for the simplicity and conciseness of the thesis. However, the interested reader can refer to our work [57] for simulation based and analytical SER plots that we derive in this chapter for the MRC receiver case. In fact, MRC receiver, which performs fairly well in terms of achievable rate with no quantization [58], suffers from an error

floor with 1-bit quantization [59] for SR sampling, which also seems to exist for the FTSR sampling case [57]. The reason to observe worse error rate and achievable rate performance with FTSR sampling compared to SR sampling for the MRC receiver case is attributed to the fact that ISI is introduced with oversampling in addition to the MUI in SR case, both of which cannot be suppressed efficiently with the MRC receiver (which won't be the case for ZF type receiver, which is able to suppress ISI and MUI much better, as to be seen later).

For the ZF type receiver, SER and achievable rate per user plots for oversampling rate $\beta = 1$ and $\beta = 2$ are presented in Fig.3.2 and Fig.3.3. In Fig.3.2, the SER curves that are obtained with Monte-Carlo based simulations for oversampling rates $\beta = 1$ and $\beta = 2$ are referred to as empirical $\beta = 1$ and $\beta = 2$, respectively. SER curves named as proposed UB (3.40) $\beta = 1$ and $\beta = 2$ correspond to the proposed SER upper bound in this chapter, which can be calculated using (3.40), for $\beta = 1$ and $\beta = 2$. Moreover, the curve with the label "Approximation in [36]" is the SER plot according to the approximate analytical expression in [36] for the SR sampling case. The plots with the label "chan. est." corresponds to the plots with BLMMSE channel estimation, which are plotted with dashed lines, whereas perfect CSI curves are plotted with solid lines.

Observed from Fig.3.2, oversampling by 2 results in about 2 dB SNR advantage compared to SR sampling case ($\beta = 1$ case) either with perfect or imperfect CSI. Moreover, mutual information per user for $\beta = 2$ case is considerably above the $\beta = 1$ case, which can be inferred from Fig.3.3. Moreover, from Fig.3.2 and Fig.3.3 it can be inferred that our analytical SER upper bound and achievable rate lower bound are very close to the simulation values for $\beta = 1$ and perfect CSI, much closer than the analytical SER and achievable rate per user curves based on the expression derived in [36] for perfect CSI. The reason for the discrepancy between the analytical SER and achievable rate per user expressions in [36] and the simulated values is owing to the fact that analytical expressions in [36] do not consider the inter-antenna correlations given x_m and $\mathbf{H}$, which we take into account by reflecting these correlations through the matrix $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ in (3.39). In fact, we prove in Proposition 1 that the correlation between the received quantized signals from different antennas when the channel matrix is given is non-zero.

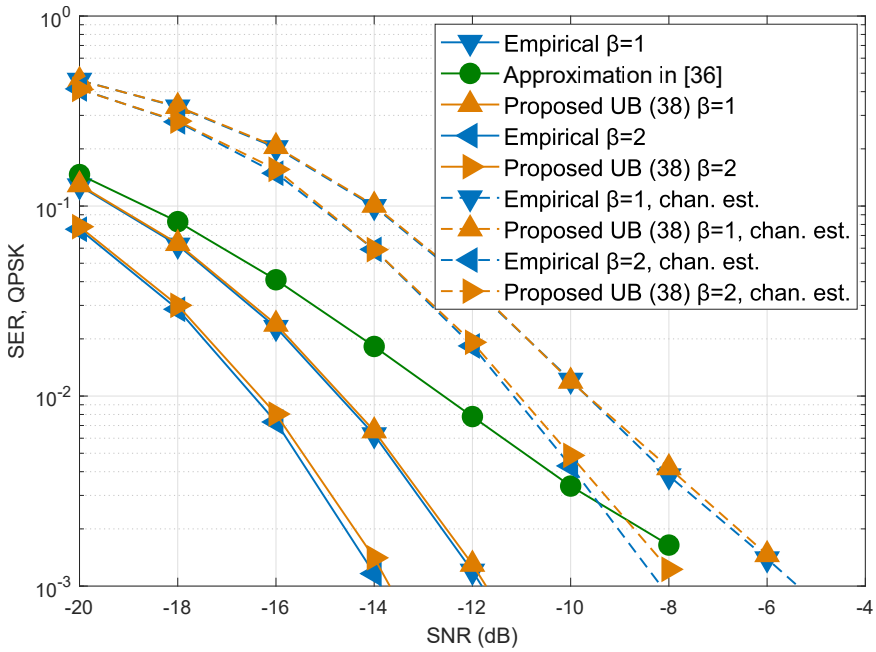


Figure 3.2: Analytical and simulation based SER vs. SNR curves for $M = 400$, $K = 20$, oversampling rate $\beta = 1, 2$ with ZF detector. ($\rho = 0.8$, QPSK modulation) $\tau = K$ for BLMMSE channel estimation.

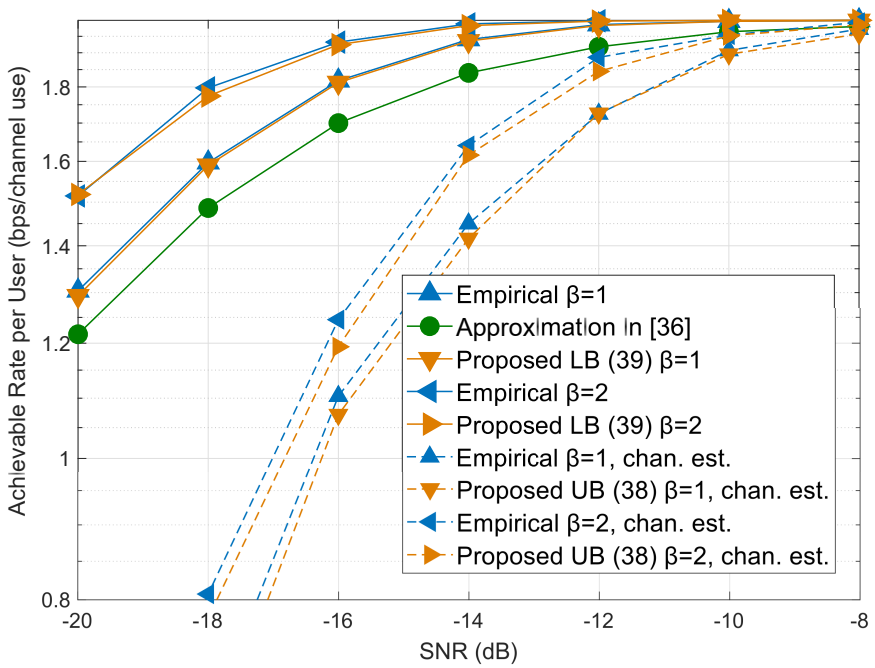


Figure 3.3: Analytical and simulation based achievable rate per user curves for $M = 400$, $K = 20$, oversampling rate $\beta = 1, 2$ with ZF detector. ($\rho = 0.8$) $\tau = K$ for BLMMSE channel estimation.

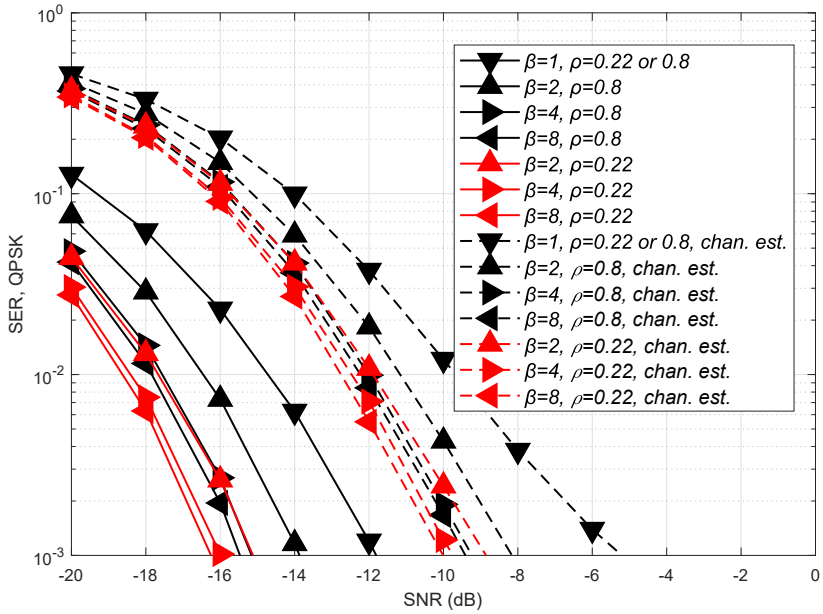


Figure 3.4: Simulation based SER vs. SNR curves for $M = 400$, $K = 20$, for $\rho = 0.22$ or 0.8, oversampling rate $\beta = 1, 2, 4, 8$ with ZF detector, QPSK modulation. $\tau = K$ for BLMMSE channel estimation.

In Fig.3.2 and Fig.3.3, the proposed upper and lower bounds are also very close to the simulated values for perfect and imperfect CSI cases for $\beta = 1$ or $\beta = 2$. In addition, it can be seen that there is a significant performance difference between the perfect and imperfect CSI cases in the aforementioned figures, which will also be the case for the subsequent simulation results. This is owing to the fact that 1-bit quantization results in a significant distortion in the channel estimates. However, this distortion can be decreased without an error-floor as long as the training length is increased, as proven in [30]. Our aim in this study is to show that the advantages with oversampling persists even when the channel estimates are of low-quality.

We also provide simulation based SER and achievable rate curves for 1-bit quantized massive MIMO system with ZF type receiver for higher oversampling rates of 4 or 8 in Fig.3.4 and Fig.3.5.

Inferred from Fig.3.4, oversampling by 8 provides about 4 dB SNR gain compared to SR sampling case for the SER value of 10^{-3} when roll-off factor $\rho = 0.8$ for perfect or imperfect CSI cases. This SNR gain is up to 5 dB when $\rho = 0.22$, which is the roll-off factor specified for square RRC filter in UMTS [60]. The reason to

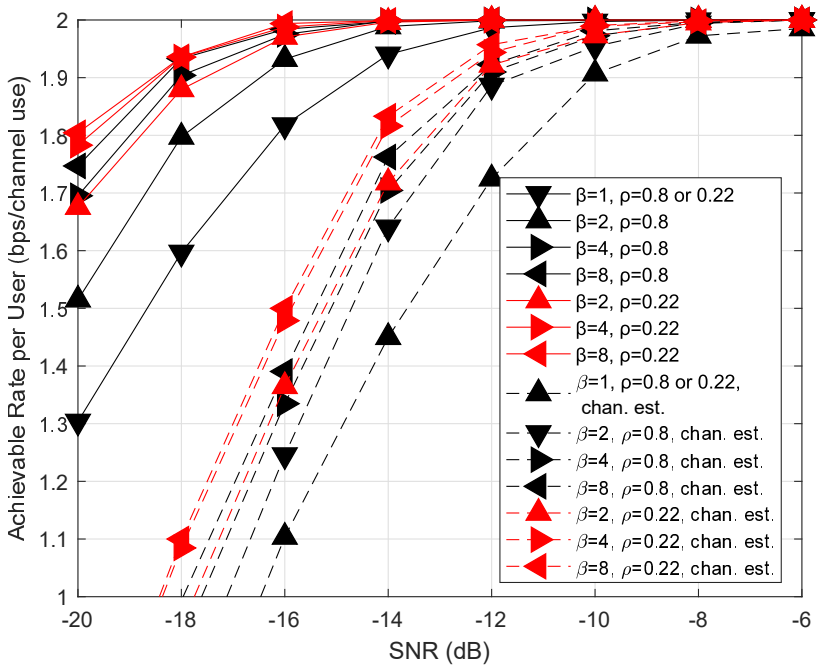


Figure 3.5: Simulation based achievable rate per user curves for $M = 400$, $K = 20$, for $\rho = 0.22$ or 0.8, oversampling rate $\beta = 1, 2, 4, 8$ with ZF detector. $\tau = K$ for BLMMSE channel estimation.

observe better SER performance when the roll-off factor is decreased is attributed to the fact that as the roll-off factor gets smaller, the transmitted pulse shape decays slower so that additional FTSR samples accumulate higher symbol energy compared to the increase in noise. Another important observation is that the same SNR gain obtained with oversampling by 4 with $\rho = 0.8$ can be achieved with oversampling only by 2 when $\rho = 0.22$ for perfect or imperfect CSI. Therefore, in terms of error-rate performances, it seems that it is better to use low roll-off factor RRC pulses with oversampling, which also reduces the excess bandwidth usage. Another important observation is that there is no significant difference between the SNR gain obtained with oversampling by 4 and 8, for both of the roll-off factor cases, thus oversampling by 4 can be considered to be enough for the investigated scenarios. Moreover, in Fig.3.5 the achievable rate per user goes above 1.997 bps/Hz at about -14.8 dB for $\beta = 8$ and $\rho = 0.22$ and perfect CSI, whereas this number is about -10 dB for $\beta = 1$, pointing out an SNR advantage of roughly 5 dB in terms of achievable rate per user. A similar SNR advantage is also observed for imperfect CSI case in Fig.3.5.

Although FTSR sampling has the aforementioned advantages, it can be stated that

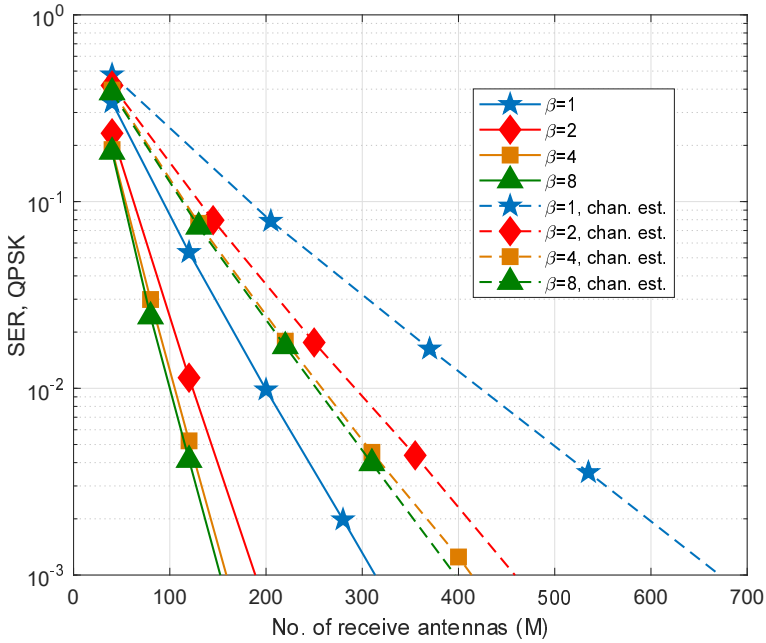


Figure 3.6: Simulation based SER vs number of receive antennas (M) for oversampling rate $\beta = 1, 2, 4, 8$ with ZF detector for SNR=-10 dB, $K = 20$. ($\rho = 0.22$, QPSK modulation) $\tau = K$ for BLMMSE channel estimation.

it causes increased signal processing complexity at the receiver side owing to an increased number of samples to be processed. However, if the advantage of oversampling is exploited as decreasing the number of receive antennas without any error rate performance degradation, the signal processing complexity at the receiver side can be maintained at feasible levels since the number of samples taken at the receiver side will be reduced owing to smaller number of receive antennas. In this case, the advantage will occur as the reduced form factor of the array to be deployed which can be critical for implementation purposes. To observe the advantage of oversampling from that viewpoint, SER of the 1-bit quantized MIMO system with ZF receiver is plotted against the number of antennas for -10 dB SNR level in Fig.3.6 for $\rho = 0.22$ and various β values.

Observed from Fig.3.6, while we can achieve the SER level of 10^{-3} for SR sampling case (for $\beta = 1$) with about 300 receive antennas for perfect CSI case, we need about 200 antennas for $\beta = 2$ to achieve the same SER level with perfect channel knowledge. This number can fall down to 150 antennas when $\beta = 8$. This clearly shows that oversampling can reduce necessary number of receive antennas significantly without

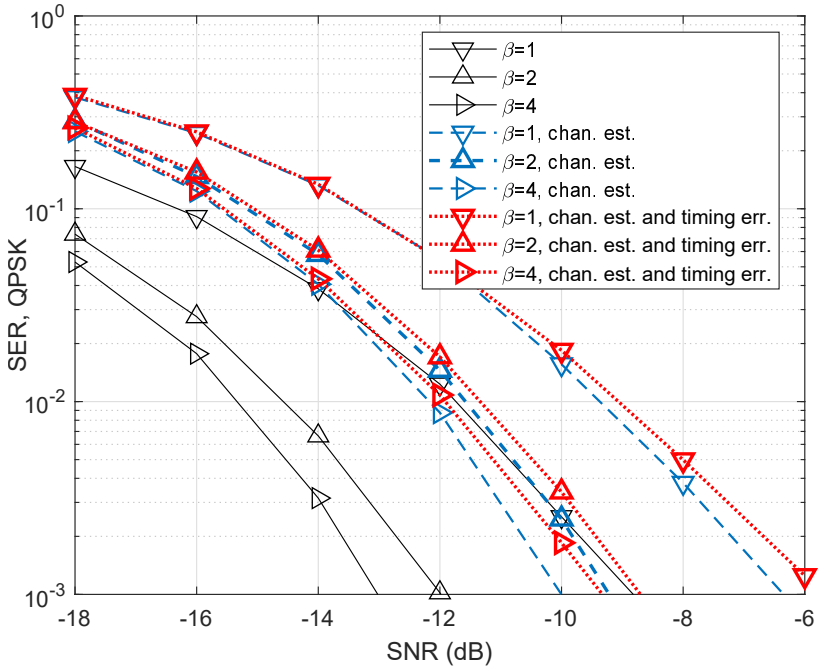


Figure 3.7: Simulation based SER vs. SNR curves with channel estimation and timing errors for $M = 200$, $K = 10$, $\rho = 0.22$, oversampling rate $\beta = 1, 2, 4$ with ZF detector, QPSK modulation. For channel estimation $\tau = 3K$ and $\sigma_e = 0.05T$.

performance degradation. A similar reduction in the necessary number of antennas is observed with oversampling for imperfect CSI case.

To present the impact of timing error, SER performance of 1-bit quantized massive MIMO system under timing and channel estimation errors is plotted in Fig.3.7 with the number of antennas $M = 200$ and $K = 10$. Perfect CSI and timing cases correspond to solid curves, while imperfect CSI and timing cases are plotted with dashed curves. The SER curves that are obtained under both imperfect CSI and timing error are labelled as “chan. est. and timing err.". As can be observed in Fig.3.7, while about 4 dB advantage is obtained with oversampling by 4 in perfect CSI and timing cases, this SNR advantage does not change much under channel estimation errors or when both channel estimation and timing error exist. Therefore, it can be stated that the SNR advantage obtained with temporal oversampling is maintained under channel estimation and timing errors.

3.7 Conclusion

In this chapter, FTSR sampling has been proposed for uplink massive MIMO systems with 1-bit quantization. Moreover, a BLMMSE channel estimation scheme has been proposed based on the BLMMSE channel estimation techniques that exist for symbol rate sampling in literature. With FTSR sampling in such systems, we have observed that we can achieve about 4-5 dB SNR advantage with the ZF receiver in terms of SER and achievable rate compared to SR sampling case both with channel estimation and timing errors and without. We have also observed that we can reduce the required number of receive antennas significantly (up to $\%50$ percent) by using FTSR sampling without any performance degradation compared to the SR sampling case with perfect CSI or with channel estimation. The finding that FTSR reduces the necessary number of receive antennas to achieve a certain error rate performance for massive MIMO arrays can be very important in terms of computational complexity and the hardware implementation that may require limited form factors for the receive antenna array.

In addition to the simulation based observations regarding the advantages of FTSR sampling in 1-bit quantized massive MIMO systems, we have also derived an upper bound on SER and a lower bound on the achievable rate for such systems for both perfect and imperfect CSI. We have also proved that the bounds we provide are tight for low SNR regime. Furthermore, we have observed that the bounds that we have derived are close to the simulation based curves. Moreover, the bounds that have been provided here are also applicable to 1-bit quantized massive MIMO systems with no FTSR sampling and predict empirical results better than the approximate analytical curves existing in literature.

In short, we take the first step in this chapter to show the benefits of oversampling in time for massive MIMO systems with low-resolution quantizers. The results are quite promising, leading us to analyze such systems with oversampling for frequency-selective fading channels in the next chapter.

CHAPTER 4

UPLINK PERFORMANCE ANALYSIS OF OVERSAMPLED WIDEBAND MASSIVE SC-MIMO WITH ONE-BIT ADCS

4.1 Motivation and Contributions

In this chapter, we extend the receiver in Chapter 3 that works with samples taken faster than symbol rate such that it can work under frequency-selective fading channels with single-carrier modulation. As frequency-selective fading is present for most of the practical channels, investigation of temporal oversampling in frequency-selective channels is critical. The related works to the content of this chapter are the same as the ones mentioned in the related works of the study in Chapter 3, thus will not be repeated.

For the extension to frequency-selective case, we begin with constructing the signal model for temporally oversampled wideband channels and formulate the ZF detector accordingly. Then, the performance analysis in terms of SER and achievable rate of oversampled wideband massive single-carrier MIMO (SC-MIMO) with one-bit ADCs and ZF detectors are presented. Similar to the flat fading case, analytical bounds are derived for SER and achievable rate for wideband fading channel. The analytical bounds are compared to the simulated values in the section devoted to the simulation results. The analysis and the simulations convey that even more significant performance gains can be obtained with temporal oversampling for frequency-selective channels compared to the flat fading channel case. To sum up, the main contribution items associated with this chapter are as follows:

- We apply temporal oversampling as a novel technique to mitigate MUI and inter-symbol interference and obtain better error rate performance for uplink

wideband massive SC-MIMO systems with 1-bit ADCs. By expressing the input-output relation for such systems in a simple form as in (4.11), we derive the ZF detector for oversampled case, which has been shown to perform as good as an ML receiver for 1-bit quantized MIMO structures [53] with a large number of receive antennas for the SR sampling case. By employing ZF receiver with oversampling, we achieve up to 9 dB SNR gain compared to SR sampling under perfect and imperfect CSI cases. Moreover, we also show that the necessary number of antennas to achieve a certain error rate performance can be lowered significantly (up to 70% reduction) with temporal oversampling.

- We make the error rate performance analysis for wideband 1-bit quantized uplink massive MIMO structures with temporal oversampling at the receiver side for both perfect and imperfect CSI cases. The accuracy of the analysis is verified by the simulation based results.

4.2 Signal Model

We start from the received and pulse-matched filtered signal at the mth antenna expressed for SC-MIMO in (2.4) in Chapter 2, which can be rewritten as

$$d_m(t) = \sum_{\ell=0}^{L-1} \sum_{k=1}^{K} \sum_{n=1}^{N} h_{m,k}[\ell] x_{n,k} p(t-(n-1)T-\ell T) + z_m(t), \tag{4.1}$$

where $h_{m,k}[\ell]$ is the ℓth channel tap between the mth antenna and the kth user, $x_{n,k}$ is the transmitted data symbol of user k at the nth symbol period, $z_m(t)$ and $p(t)$ is the matched filtered pulse shape and noise, as defined in Chapter 2. When $p(t)$ is a Nyquist pulse, symbol-rate sampling of $d_m(t)$ results in a discrete-time signal model consistent with that in [23], which is one of the fundamental works analyzing quantized wideband one-bit massive MIMO. We define vector $\mathbf{y}$ as

$$\mathbf{y} = \begin{bmatrix} [\mathbf{y}^{\mathbf{SR}}]^T & [\mathbf{y}^{\mathbf{OS},1}]^T & [\mathbf{y}^{\mathbf{OS},2}]^T & \cdots & [\mathbf{y}^{\mathbf{OS},\beta-1}]^T \end{bmatrix}^T_{1\times\beta MN}, \tag{4.2}$$

where

$$\mathbf{y}^{\mathbf{SR}} = \begin{bmatrix} y^{SR}_{1,1} & y^{SR}_{1,2} & \cdots & y^{SR}_{1,M} & y^{SR}_{2,1} & \cdots & y^{SR}_{N,M} \end{bmatrix}^T_{1\times MN}, \tag{4.3}$$

$$\mathbf{y}^{\mathbf{OS},b} = \begin{bmatrix} y_{1,1}^{OS,b} & \cdots & y_{1,M}^{OS,b} & y_{2,1}^{OS,b} & \cdots & y_{N,M}^{OS,b} \end{bmatrix}_{1\times MN}^{T}, \tag{4.4}$$

$b = 1, ..., \beta - 1$ with positive integer β. In (3.3) and (3.4),

$$y_{i,m}^{SR} = d_m((i-1)T), \tag{4.5}$$

which corresponds to the samples taken at the symbol rate and

$$y_{i,m}^{OS,b} = d_m((i-1)T + bT/\beta), \tag{4.6}$$

corresponding to the FTSR samples. Furthermore, we also define vectors $\mathbf{x}$ and $\mathbf{n}$ as

$$\mathbf{x} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,K} & x_{2,1} & \cdots & x_{N,K} \end{bmatrix}_{1\times NK}^{T}, \tag{4.7}$$

$$\mathbf{n} = \begin{bmatrix} [\mathbf{n}^{\mathbf{SR}}]^T & [\mathbf{n}^{\mathbf{OS},1}]^T & [\mathbf{n}^{\mathbf{OS},2}]^T & \cdots & [\mathbf{n}^{\mathbf{OS},\beta-1}]^T \end{bmatrix}_{1\times \beta MN}^{T}, \tag{4.8}$$

where

$$\mathbf{n}^{\mathbf{SR}} = \begin{bmatrix} n_{1,1}^{SR} & n_{1,2}^{SR} & \cdots & n_{1,M}^{SR} & n_{2,1}^{SR} & \cdots & n_{N,M}^{SR} \end{bmatrix}_{1\times MN}^{T}, \tag{4.9}$$

$$\mathbf{n}^{\mathbf{OS},b} = \begin{bmatrix} n_{1,1}^{OS,b} & \cdots & n_{1,M}^{OS,b} & n_{2,1}^{OS,b} & \cdots & n_{N,M}^{OS,b} \end{bmatrix}_{1\times MN}^{T}, \tag{4.10}$$

$b = 1, ..., \beta - 1$, In (4.9) and (4.10), $n_{i,m}^{SR} = z_m((i-1)T)$ and $n_{i,m}^{OS,b} = z_m((i-1)T + bT/\beta)$. In this case, (4.1) can be written in matrix-vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.11}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_{\beta-1} \end{bmatrix}^{T}, \tag{4.12}$$

in which

$$\mathbf{G}_b = \begin{bmatrix} \mathbf{\Gamma}_1^b & \mathbf{\Gamma}_0^b & \mathbf{\Gamma}_{-1}^b & \cdots & \mathbf{\Gamma}_{-(N-2)}^b \\ \mathbf{\Gamma}_2^b & \mathbf{\Gamma}_1^b & \mathbf{\Gamma}_0^b & \cdots & \mathbf{\Gamma}_{-(N-1)}^b \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{\Gamma}_N^b & \mathbf{\Gamma}_{N-1}^b & \cdots & \mathbf{\Gamma}_2^b & \mathbf{\Gamma}_1^b \end{bmatrix}_{MN\times NK}^{T}, \tag{4.13}$$

where $b = 1, ..., \beta - 1$, $\mathbf{\Gamma}_i^b = \sum_{\ell=0}^{L-1} \gamma_{i,\ell}^b \mathbf{C}_\ell$, $\gamma_{i,\ell}^b = p((i-1)T + bT/\beta - \ell T)$ and $\mathbf{C}_\ell$ is the matrix, whose element at its mth row and the kth column is equal to $h_{m,k}[\ell]$.

In the case that 1-bit quantized version of the received signal vector $\mathbf{y}$ is taken into account, the model in (4.11) becomes

$$\mathbf{r} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}(\mathbf{H}\mathbf{x} + \mathbf{n}), \tag{4.14}$$

where $Q(\mathbf{y})$ =sgn$(\mathrm{Re}(\mathbf{y})) + jsgn(\mathrm{Im}(\mathbf{y}))$, sgn(.) being the signum function, $\mathrm{Re}(.)$ and $\mathrm{Im}(.)$ takes the real and imaginary parts of their operands. Without loss of generality, we assume ZF type linear receiver. In that case, the soft estimate for the transmitted symbol vector $\mathbf{x}'$ can be found as [36, 53]

$$\mathbf{x}' = \mathbf{B}\mathbf{r}, \tag{4.15}$$

where $\mathbf{B}$ is a linear receive filter. For MRC and ZF receiver, $\mathbf{B} = \hat{\mathbf{H}}^H$ or $\mathbf{B} = (\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^H$, respectively, where $\hat{\mathbf{H}}$ is the estimate of the channel matrix $\hat{\mathbf{H}}$. The hard symbol estimate vector $\hat{\mathbf{x}}$ for the transmitted symbol vector $\mathbf{x}$ is found by mapping the elements of the soft estimate vector $\mathbf{x}'$ to the minimum distance constellation point.

In this chapter, we prefer leaving the proposal of a channel estimation algorithm to a future work, yet we consider the effect of imperfect CSI as follows. We assume that the ML estimates for the channel matrices $\mathbf{C}_\ell$ are estimated somehow. Since ML estimates are asymptotically Gaussian and unbiased, we construct the estimates for the $\mathbf{C}_l$ matrices by adding independent zero mean (since the estimates are unbiased) complex Gaussian random variables with variance σ_h^2/L to each element of the matrix $\mathbf{C}_\ell$, to form the matrix $\hat{\mathbf{C}}_\ell$. Then, $\hat{\mathbf{H}}$ can be obtained from $\hat{\mathbf{C}}_\ell$ using (4.12) and (4.13). The reason for division by the number of channel taps L for the variance of the Gaussian random variables added to each element of $\mathbf{C}_\ell$ is to normalize the received signal power from each user.

4.3 Error Rate Analysis

In this section, we make the error rate analysis for 1-bit quantized uplink wideband massive SC-MIMO with temporal oversampling at the receiver side. SER for the pth

element of the transmitted data vector $\mathbf{x}$, namely x_p, can be written as

$$\text{SER} = \mathbf{E}_{\mathbf{H}} \left[\sum_{x_p} \sum_{\hat{x}_p \neq x_p} p(\hat{x}_p | x_p, \mathbf{H}) p(x_p) \right], \tag{4.16}$$

where $\hat{x}_p$ is the hard symbol estimate of the transmitted symbol x_p and $p(x_p)$ is the probability mass function of x_p. The index p is selected to be such that it corresponds to one of the symbols that is transmitted around the center of the transmitted data block. In this way, SER that considers all elements of vector $\mathbf{x}$, except the ones that are transmitted at the beginning and the end of the transmitted data block, will be similar to the SER of x_p since the channel coefficients for different symbol intervals and users have identical distributions. However, the error rate for the symbols at the block edges may be different than the symbols that are transmitted around the center of the block since the number of FTSR samples that refine the estimates at the block edges will be more limited as compared to the ones that are around the center of the transmitted data block. Considering the structure of $\mathbf{x}$ defined in (4.7), if x_p is the pth element of $\mathbf{x}$, it belongs to the mth user, where $m = \text{mod}\ (p-1, K) + 1$ and the nth symbol interval, where $n = \lfloor \frac{p-1}{K} \rfloor + 1$. Here $\lfloor\ .\ \rfloor$ is the floor function that gives the largest integer less than its operand and mod (a, p) yields the remainder after division of a by p.

To be able to calculate SER using (4.16), $p(\hat{x}_p | x_p, \mathbf{H})$ needs to be found. In order to reach a a tractable analysis, we will show that $p(\hat{x}_p | x_p, \mathbf{H})$ can be approximated well with a normal distribution. In such a case, $p(\hat{x}_p | x_p, \mathbf{H})$ will be uniquely defined when its mean and variance is found.

To start with, considering (4.11), the mth element of $\mathbf{y}$, namely y_m, can be expressed as

$$y_m = \sum_{k=1}^{NK} h_{m,k} x_k + n_m, \tag{4.17}$$

where x_k and n_m are the kth and m^{th} element of transmitted data vector $\mathbf{x}$ and $\mathbf{n}$ and $h_{m,k}$ is the element of $\mathbf{H}$ at the mth row and k^{th} column. Define $z_m \triangleq \sum_{k=1}^{NK} h_{m,k} x_k$. Since all x_k's are independent, which makes $h_{m,k} x_k$ terms in the z_m expression independent conditioned on $\mathbf{H}$ and x_p, and each $h_{m,k} x_k$ term has a finite variance, the distribution of the random variable z_m conditioned on x_p and $\mathbf{H}$ converges to that of a Gaussian by the CLT [54]. This implies that $y_m = z_m + n_m$ will be Gaussian since

n_m is Gaussian. The conditional mean of y_m can be found as

$$\mathbf{E}[y_m|x_p, \mathbf{H}] = h_{m,p}x_p \tag{4.18}$$

since $\mathbf{E}[x_k] = 0$ for $k \neq p$ and $\mathbf{E}[n_m] = 0$. The conditional variance of y_m can also be written as

$$Var[y_m|x_p, \mathbf{H}] = \sum_{k \neq p} |h_{m,k}|^2 + \sigma_n^2 \tag{4.19}$$

since $\mathbf{E}[|x_k|^2] = 1$ for $k \neq p$. Therefore, it can be stated that the probability density function of y_m conditioned on x_p and $\mathbf{H}$, namely $f(y_m|x_p, \mathbf{H})$, satisfies

$$f(y_m|x_p, \mathbf{H}) \approx \mathcal{CN}(h_{m,p}x_p, \sum_{k \neq p} |h_{m,k}|^2 + \sigma_n^2). \tag{4.20}$$

The observation after 1-bit quantization, namely $r_m = \mathcal{Q}(y_m)$, corresponds to the mth element of the vector $\mathbf{r}$ defined in (4.14). The conditional mean of r_m given x_p and $\mathbf{H}$, which is denoted by $\mu_{m|p}$ can be expressed as

$$\mu_{m|p} = (1+j)P_{1m|p} + (-1+j)P_{2m|p} + (1-j)P_{3m|p} + (-1-j)P_{4m|p}, \tag{4.21}$$

where $P_{1m|p} = Pr(r_m = 1+j|x_p, \mathbf{H})$. Similarly, $P_{2m|p} = Pr(r_m = -1+j|x_p, \mathbf{H})$, $P_{3m|p} = Pr(r_m = 1-j|x_p, \mathbf{H})$, $P_{4m|p} = Pr(r_m = -1-j|x_p, \mathbf{H})$. Due to (4.20), these probabilities can be calculated using the standard error function. The point where $\mu_{m|p}$ values will be used will be clear shortly.

Due to (4.15), the pth element of the soft symbol estimate vector $\mathbf{x}'$, which can be denoted by x'_p, can be written as

$$x'_p = \mathbf{b}_p^T \mathbf{r}, \tag{4.22}$$

where $\mathbf{b}_p^T$ is the row vector equal to the pth row of the linear receive filter matrix $\mathbf{B}$ in (4.15). (4.22) can be reexpressed as

$$x'_p = \sum_{m=1}^{MN\beta} b_{p,m} r_m, \tag{4.23}$$

where $b_{p,m}$ is the element of receive filter matrix $\mathbf{B}$ on the pth row and m^{th} column. Considering the expression for x'_p, it is composed of a summation of many terms of finite variance. Therefore, relying on the Central Limit Theorem, we make the assumption that x'_m approximately has normal distribution. This assumption is verified

with empirical observations in [36] for symbol-rate sampling case. For oversampled case, the Gaussian assumption has better accuracy as the number of independent observations in the calculation of x'_p increase for higher β. Therefore, it can be stated that

$$f(x'_p|x_p, \mathbf{H}) \approx \mathcal{CN}(\sum_{m=1}^{MN\beta} b_{p,m}\mu_{m|p}, \sigma^2_{x'_p|x_p}), \tag{4.24}$$

where $\sigma^2_{x'_p|x_p}$ is the conditional variance of x'_p conditioned on x_p. Since $b_{p,m}$ is known and $\mu_{m|p}$ can be found from (4.21), the conditional mean of x'_p given x_p and $\mathbf{H}$ whose expression is given in (4.24) can be calculated. What remains is to find $\sigma^2_{x'_p|x_p}$. Considering (4.23), $\sigma^2_{x'_p|x_p}$ can be calculated by using the variance of sum formula [55] as

$$\sigma^2_{x'_p|x_p} = \mathbf{b}_p^T \mathbf{C}_{\mathbf{rr}} \mathbf{b}_p^*, \tag{4.25}$$

where $\mathbf{C}_{\mathbf{rr}} = \mathbf{E}[(\mathbf{r} - \mathbf{E}[\mathbf{r}])(\mathbf{r} - \mathbf{E}[\mathbf{r}])^H|x_p, \mathbf{H}]$ is the covariance matrix for vector $\mathbf{r}$. Therefore, if $\mathbf{C}_{\mathbf{rr}}$ is calculated, the value of $\sigma^2_{x'_p|x_p}$ can be found. When $\sigma^2_{x'_p|x_p}$ is found, along with the calculated mean value of x'_p given x_p and $\mathbf{H}$, it means that $f(x'_p|x_p, \mathbf{H})$ can be found according to (4.24). Once $f(x'_p|x_p, \mathbf{H})$ is known, the probability mass function for the hard symbol estimate $p(\hat{x}_p|x_p, \mathbf{H})$ can be calculated using Q-functions. Then, SER can be calculated by replacing the calculated $p(\hat{x}_p|x_p, \mathbf{H})$ values into the expression in (4.16).

To find $\mathbf{C}_{\mathbf{rr}}$, we first define the vectors $\mathbf{r}'$ and $\mathbf{y}'$ as follows.

$$\mathbf{r}' = \begin{bmatrix} \mathrm{Re}(\mathbf{r}) \\ \mathrm{Im}(\mathbf{r}) \end{bmatrix}, \mathbf{y}' = \begin{bmatrix} \mathrm{Re}(\mathbf{y}) \\ \mathrm{Im}(\mathbf{y}) \end{bmatrix}. \tag{4.26}$$

According to arcsine law, the following equation holds [56].

$$\mathbf{R}_{\mathbf{r}'\mathbf{r}'} = \frac{2}{\pi}\left[\arcsin\left(\mathrm{diag}(\mathbf{R}_{y'y'})^{-\frac{1}{2}}\mathbf{R}_{y'y'}\mathrm{diag}(\mathbf{R}_{y'y'})^{-\frac{1}{2}}\right)\right]. \tag{4.27}$$

In (4.27), $\mathbf{R}_{\mathbf{r}'\mathbf{r}'} = \mathbf{E}[\mathbf{r}'\mathbf{r}'^H]$ and $\mathbf{R}_{\mathbf{y}'\mathbf{y}'} = \mathbf{E}[\mathbf{y}'\mathbf{y}'^H]$. However, (4.27) is valid only when $\mathbf{r}'$ is zero mean and $\mathbf{y}'$ is Gaussian. It has been discussed that $\mathbf{y}'$ is approximated as Gaussian as in (4.20). However, $\mathbf{r}'$ is not zero mean. Therefore, we will find a value for $\sigma^2_{x'_p|x_p}$ by assuming $x_p = 0$, which implies that $\mathbf{r}'$ is also zero mean, thus (4.27) is valid. It can be shown that the value of $\sigma^2_{x'_p|x_p}$ found under this assumption is larger than the actual $\sigma^2_{x'_p|x_p}$, for which x_p is non-zero. If this value is used in place of $\sigma^2_{x'_p|x_p}$, to find $f(x'_p|x_p, \mathbf{H})$ and $p(\hat{x}_p|x_p, \mathbf{H})$ and the SER is found using (4.16), the calculated

SER will be higher than the actual value of SER, thus will constitute an upper bound (UB) for SER. The proof for the validity of this upper bound can be made similar to the proof made for frequency-flat fading channel case investigated in Chapter 3. Moreover, the details on how to find $\mathbf{R_{rr}}$ from $\mathbf{R_{r'r'}}$ and $\mathbf{R_{y'y'}}$ based on $\mathbf{H}$, data and noise correlation matrices $\mathbf{R_{xx}}$ and $\mathbf{R}_{nn}$ is also similar to finding $\mathbf{R_{rr}}$ and $\mathbf{R_{\tilde{y}\tilde{y}}}$ in Chapter 3, whose details are provided in Appendix A.2.

4.4 Simulation Results

Number of users (K) and receive antennas (M) are taken to be 20 and 400, respectively. The block length (N) is selected to be 10 symbols and the oversampling rate (β) is chosen between 1 and 8. The channel length L is chosen to be either $L = 3$ or $L = 10$. The power-delay profile of the channel is taken as uniform, that is $\rho_k[\ell] = 1/L$ for $\ell = 0, 1, ..., L-1$ and $k = 1, 2, ..., K$. The simulation based SER plots are obtained by taking the average of 100 channel matrix, noise and symbol vector triplets. For the analytical SER plots, we find $p(\hat{x}_p|x_p, \mathbf{H})$ as decribed in Section 4.3 and find the analytical SER using (4.16) for a certain channel matrix realization and the averaging with respect to channel matrix is performed over 100 channel matrix realizations. Note that the calculated analytical SER is in fact an upper bound for the actual SER as mentioned in Section 4.3. The analytical SER is found for the data symbol that is transmitted in the 5th symbol interval (which is around the middle of the transmitted block of length 10) by the 20th user. Considering the structure of vector $\mathbf{x}$ in (4.7), this means that $x_p = x_{100}$. The SNR values on the plots are equal to $1/\sigma_n^2$.

The SER performance for quantized massive MIMO system with ZF receiver is plotted for $\beta = 1$ and $\beta = 2$ in Fig. 4.1 when $L = 3$ and $\rho = 0.8$ for both perfect and imperfect CSI cases. In Fig. 4.1, the SER curves that are obtained with Monte-Carlo based simulations for oversampling rates $\beta = 1$ and $\beta = 2$ are referred to as $\beta = 1$ and $\beta = 2$ empirical, respectively. SER curves referred to as “proposed UB $\beta = 1$ and $\beta = 2$" correspond to the proposed SER UB in this chapter for $\beta = 1$ and $\beta = 2$, respectively. The dashed curves with the label “Imperf CSI" are the performance curves that are obtained under channel estimation errors (for $\sigma_h^2 = 0.4$), whereas the

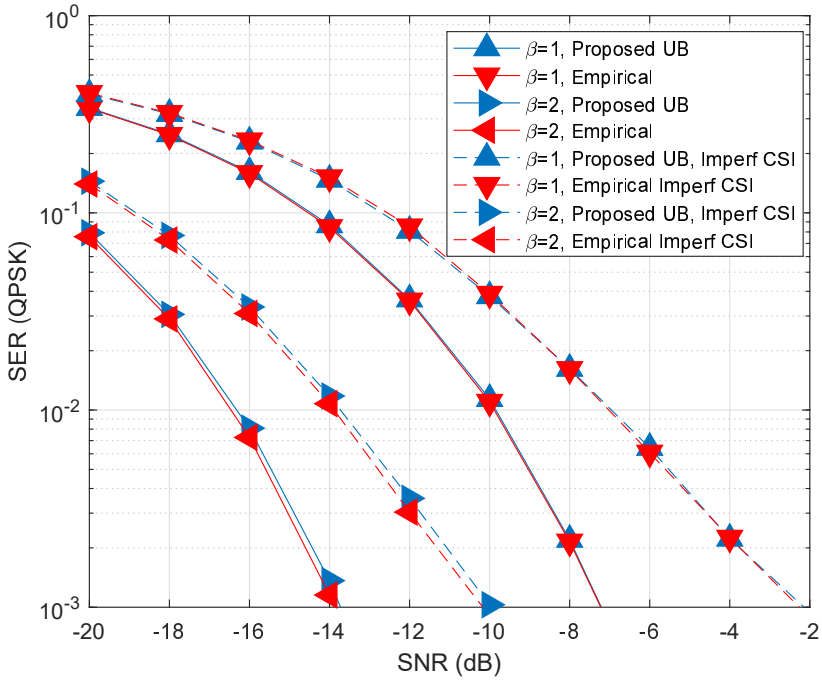


Figure 4.1: Analytical and simulation based SER vs. SNR curves for $M = 400$, $K = 20$, oversampling rate $\beta = 1, 2$, channel length $L = 3$, roll-off factor $\rho = 0.8$ with ZF detector for perfect and imperfect CSI ($\sigma_h^2 = 0.4$).

solid curves correspond to perfect CSI cases. As can be noted in Fig. 4.1, oversampling by 2 provides about 7 dB SNR gain compared to the SR sampling case when the SNR values required to attain the SER of 10^{-3} are considered. This SNR gain is not reduced under channel estimation errors. Moreover, it can also be seen that the analytical SER curves are very close to the empirical SER curves for all cases.

We also provide simulation based SER curves for higher oversampling rates of 4 or 8, various roll-off factors (for $\rho = 0.22$ and $\rho = 0.8$) in Fig. 4.2 for both perfect and imperfect CSI.

The solid curves in Fig. 4.2 correspond to perfect CSI cases whereas the dashed curves correspond to imperfect CSI cases. The blue curves are for roll-off factor $\rho = 0.8$ and the red curves are for $\rho = 0.22$. As can be noted from Fig. 4.2, oversampling by 8 provides about 8.5 dB SNR gain compared to SR sampling case for $\rho = 0.8$ when the SNR values to maintain a SER of 10^{-3} are considered. When roll-off factor ρ is decreased to 0.22, this SNR gain becomes about 9 dB. The reason to observe better SER performance when the roll-off factor is decreased is attributed to the fact that

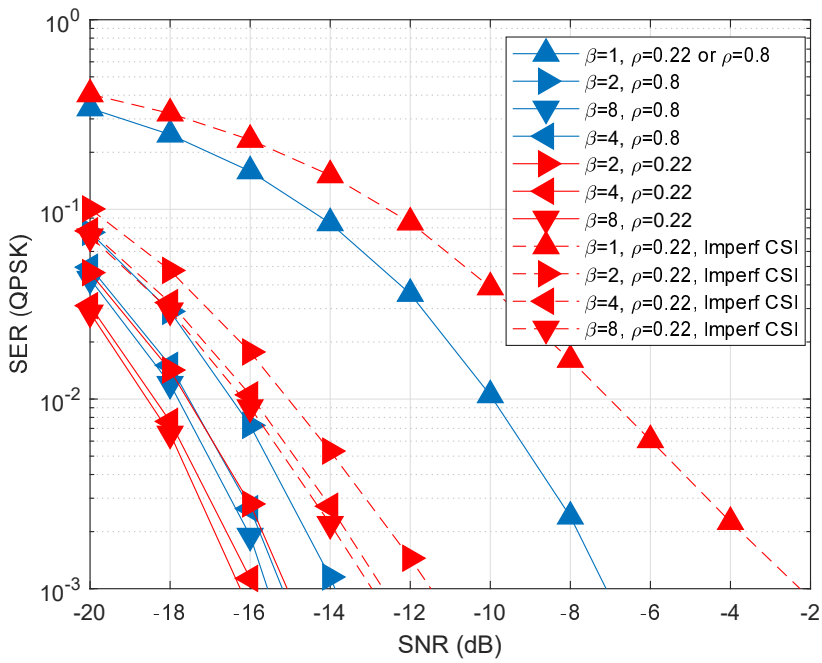


Figure 4.2: Simulation based SER vs. SNR curves for $M = 400$, $K = 20$, oversampling rate $\beta = 1, 2, 4, 8$ with ZF detector for perfect and imperfect CSI ($\sigma_h^2 = 0.4$).

as the roll-off factor gets smaller, the transmitted pulse shape decays slower so that additional FTSR samples accumulate higher symbol energy compared to the increase in noise. The mentioned SNR gains does not diminish for the imperfect CSI case. For the simplicity and conciseness of the thesis, we did not include the plots for the case $L = 10$, but we state that the SNR gains for $L = 10$ is again up to 9 dB for $\rho = 0.22$ both for perfect or imperfect CSI cases.

We also present SER versus the number of receive antennas when SNR is fixed as -7 dB for $L = 3$ and $\rho = 0.22$ in Fig. 4.3. The solid curves in Fig. 4.3 correspond to perfect CSI, whereas the dashed curves are for imperfect CSI cases. As can be noted in Fig. 4.3, while the SER of 10^{-3} can be achieved with about 400 antennas for SR sampling case ($\beta = 1$), the same SER value can be achieved with only about 190 antennas for $\beta = 2$ or with only about 150 antennas for $\beta = 4$ or $\beta = 8$ in case of perfect CSI. Under imperfect CSI conditions, SER value of 10^{-3} can be maintained by about 650 antennas for SR sampling case, whereas the same SER value is achieved only with about 300 antennas for $\beta = 2$, or with only about 250 antennas for $\beta = 4$ or $\beta = 8$. Therefore, it can be stated that the number of antennas can be reduced significantly without performance degredation. This will reduce the form factor of

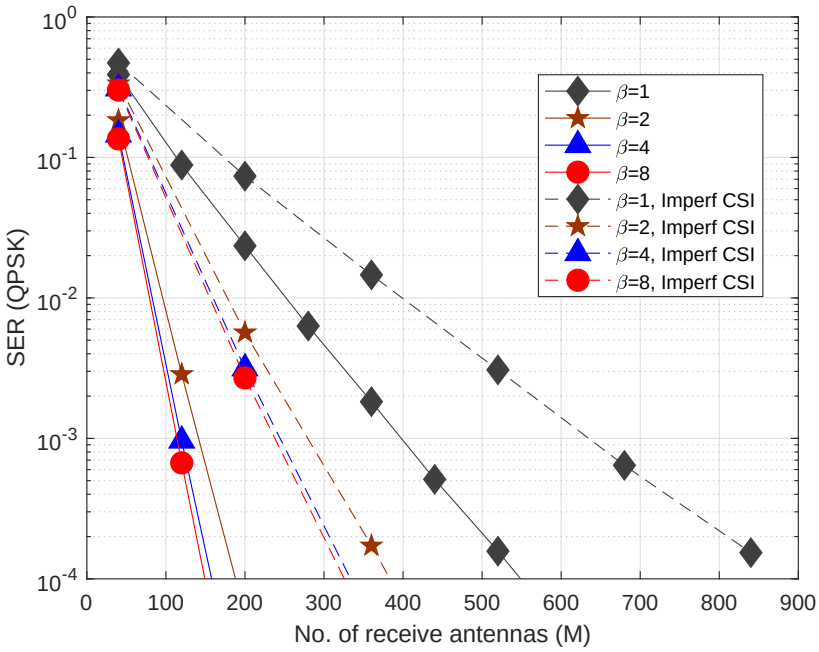


Figure 4.3: Simulation based SER vs number of receive antennas (M) for oversampling rate $\beta = 1, 2, 4, 8$ with ZF detector for SNR=-7dB under perfect and imperfect CSI ($\sigma_h^2 = 0.4$).

the antenna array, the power consumption due to RF chains owing to the fact that the number of RF chains decrease when the number of antennas are reduced.

4.5 Conclusion

In this chapter, temporal oversampling is proposed for uplink massive MIMO systems with 1-bit quantization and frequency selective channels for both perfect and imperfect CSI cases. For such systems, we have derived the ZF receiver and showed that temporal oversampling yields up to 9 dB SNR advantage compared to the SR sampling case. This is much higher than up to 5 dB SNR gain with oversampling observed for frequency-flat channels. The reason for higher gain in frequency-selective channels is that the severe quantization caused by one-bit ADC results in a very inferior performance when both ISI and MUI exist compared to the flat-fading case where only MUI is present. However, as quantization noise is suppressed with oversampling, the resulting increased ability to cancel both ISI and MUI (for frequency-

selective channel) provides a much better performance gain compared to any performance improvement attained by the increased ability to cancel only MUI in frequency-flat fading case.

Moreover, we have also made a performance analysis for such systems both under perfect and imperfect CSI. The analytically calculated values for the error rate are observed to be close to simulated values. Moreover, we have also observed that the required number of receive antennas to maintain a certain error rate performance can be reduced significantly by temporal oversampling, which in turn reduces the necessary form factor of the antenna array and the total power consumption.

In short, we have demonstrated in Chapters 3-4 that temporal oversampling has significant benefits for one-bit quantized massive MIMO for both frequency-flat and frequency-selective channels under perfect or imperfect CSI. What remains is to show that these advantages can be obtained with a detector of feasible complexity, rather than a very high complexity ZF detector that has been considered in Chapters 3-4. Such a detector is proposed in the next chapter.

CHAPTER 5

SEQUENTIAL LINEAR DETECTION IN ONE-BIT QUANTIZED UPLINK MASSIVE SC-MIMO WITH OVERSAMPLING

In previous chapters, temporal oversampling has been shown to provide significant advantages in terms of error rate performance for one-bit quantized massive single-carrier MIMO (SC-MIMO) systems. However, such an advantage is observed with a zero-forcing type receiver, whose complexity is increasing with N^3, N being the block length, which is defined as the number of data symbols that are processed in a block, making its implementation for long block lengths not possible. In this chapter, we propose a low complexity receiver for one-bit quantized uplink massive SC-MIMO whose complexity increases linearly with block length. At the same time, the SNR gains provided through temporal oversampling with the high complexity ZF receiver will be shown to be preserved with the proposed low complexity receiver. Moreover, due to the sequential structure of the proposed receiver, the delay to estimate the transmitted data symbols can be reduced to L_pT from NT, where T is the symbol duration and L_p is equal to the length of the employed pulse shape, which is much less than N in general.

5.1 Signal Model

We start from the received and pulse-matched filtered signal at the mth antenna of a massive SC-MIMO system expressed in (2.4) in Chapter 2. This signal can be written for the case of flat fading channel as

$$d_m(t) = \sum_{k=1}^{K} \sum_{n=1}^{N} c_{m,k} x_{n,k} p(t-(n-1)T) + z_m(t), \tag{5.1}$$

where $c_{m,k} = h_{m,k}[0]$ is the channel coefficient between the mth antenna and the kth user, $x_{n,k}$ is the transmitted data symbol of user k at the nth symbol period, $z_m(t)$ and $p(t)$ is the matched filtered pulse shape and noise, as defined in Chapter 2. For demonstration purposes, we define vector $\mathbf{y}$ as

$$\mathbf{y} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,M} & y_{2,1} & \cdots & y_{\beta N,M} \end{bmatrix}^T_{1\times\beta MN}, \tag{5.2}$$

where

$$y_{i,m} = d_m((i-1)T/\beta), \tag{5.3}$$

$i = 1, ..., \beta N$, $m = 1, ..., M$, β being a positive integer oversampling rate, which is defined as the ratio of the total number of samples to the samples taken at symbol rate. Furthermore, vectors $\mathbf{x}$ and $\mathbf{w}$ are also defined as

$$\mathbf{x} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,K} & x_{2,1} & \cdots & x_{N,K} \end{bmatrix}^T_{1\times NK}, \tag{5.4}$$

$$\mathbf{w} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,M} & w_{2,1} & \cdots & w_{\beta N,M} \end{bmatrix}^T_{1\times\beta MN}, \tag{5.5}$$

where $w_{i,m} = z_m((i-1)T/\beta)$, $i = 1, ..., \beta N$, $m = 1, ..., M$. In this case, (5.1) can be written in matrix-vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{5.6}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{G}_1^T & \mathbf{G}_2^T & \mathbf{G}_3^T & \cdots & \mathbf{G}_N^T \end{bmatrix}^T, \tag{5.7}$$

$$\mathbf{G}_n = \begin{bmatrix} & & \mathbf{C}_n & & \\ \gamma_n^1\mathbf{C} & \gamma_{n-1}^1\mathbf{C} & \gamma_{n-2}^1\mathbf{C} & \cdots & \gamma_{n-N+1}^1\mathbf{C} \\ \gamma_n^2\mathbf{C} & \gamma_{n-1}^2\mathbf{C} & \gamma_{n-2}^2\mathbf{C} & \cdots & \gamma_{n-N+1}^2\mathbf{C} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_n^{\beta-1}\mathbf{C} & \gamma_{n-1}^{\beta-1}\mathbf{C} & \gamma_{n-2}^{\beta-1}\mathbf{C} & \cdots & \gamma_{n-N+1}^{\beta-1}\mathbf{C} \end{bmatrix}_{\beta M\times NK}, \tag{5.8}$$

$$\mathbf{C}_n = \begin{bmatrix} \mathbf{0}_1 & \mathbf{0}_2 & \cdots & \mathbf{0}_{n-1} & \mathbf{C} & \mathbf{0}_n & \mathbf{0}_{n+1}\cdots & \mathbf{0}_{N-1} \end{bmatrix}_{M\times NK}, \tag{5.9}$$

$n = 1, ..., N$, $\gamma_n^b = p((n-1)T + bT/\beta)$ and $\mathbf{0}_n$ is a zero matrix of size $M \times K$. Moreover, $\mathbf{C}$ is the matrix, whose element at its mth row and the kth column is equal

to the channel coefficient $c_{m,k}$. The current form of $\mathbf{C}_n$ may be confusing for the cases when $n = 1$ and $n = N$, thus we specify $\mathbf{C}_1$ and $\mathbf{C}_N$ as

$$\mathbf{C}_1 = \underset{M\times NK}{\begin{bmatrix} \mathbf{C} & \mathbf{0}_1 & \mathbf{0}_2 & \cdots & \mathbf{0}_{N-1} \end{bmatrix}}, \tag{5.10}$$

$$\mathbf{C}_N = \underset{M\times NK}{\begin{bmatrix} \mathbf{0}_1 & \mathbf{0}_2 & \cdots & \mathbf{0}_{N-1} & \mathbf{C} \end{bmatrix}}. \tag{5.11}$$

As can be inferred from (5.8), each $\mathbf{G}_n$ matrix can be considered to have two parts, namely the upper part composed of the matrix $\mathbf{C}_n$ and the lower part that is consisting of γ_n^b coefficients weighting the channel matrix $\mathbf{C}$, $b = 1, ..., \beta - 1$. The part of $\mathbf{G}_n$ composed of $\mathbf{C}_n$ determines the relation between the transmitted symbols and the SR samples, whereas the remaining parts of $\mathbf{G}_n$ establishes the relation between the transmitted symbols and additional samples taken between the SR samples due to oversampling. Note that $\mathbf{C}_n$ is composed of a single matrix $\mathbf{C}$ and $N - 1$ zero matrices, which assumes that there is no ISI between the SR samples, which is a valid assumption for the narrowband channel case and when zero ISI pulse shapes are employed.

Under 1-bit quantization of the received signal vector $\mathbf{y}$, the signal model in (5.6) becomes

$$\mathbf{r} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}(\mathbf{H}\mathbf{x} + \mathbf{w}), \tag{5.12}$$

where $Q(\mathbf{y})$ =sgn$(\mathrm{Re}(\mathbf{y}))+jsgn(\mathrm{Im}(\mathbf{y}))$, sgn(.) being the signum function. The soft estimate for the transmitted symbol vector $\mathbf{x}'$ can be found as [36, 53]

$$\mathbf{x}' = \mathbf{B}\mathbf{r}, \tag{5.13}$$

where $\mathbf{B}$ is the linear receive filter matrix. MRC and ZF type receivers are given as $\mathbf{B} = \mathbf{H}^H$ and $\mathbf{B} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$, respectively. The hard symbol estimate vector $\hat{\mathbf{x}}$ is found by mapping the elements of the soft estimate vector $\mathbf{x}'$ to the minimum distance constellation point.

In this chapter, we prefer leaving the proposal of a channel estimation algorithm as a future work. However, we take into account the impact of imperfect CSI as follows. We presume that the ML estimates for the channel matrix $\mathbf{C}$ are estimated with some method. Owing to the property of the ML estimates being asymptotically Gaussian

and unbiased [61], we obtain the estimated channel matrix $\mathbf{C}$, namely $\hat{\mathbf{C}}$, by adding independent zero mean (since the estimates are unbiased) complex Gaussian random variables with variance σ_h^2 to each element of the matrix $\mathbf{C}$. Then, $\hat{\mathbf{H}}$ can be obtained from $\hat{\mathbf{C}}$ using (5.7) and (5.8). After that, $\hat{\mathbf{H}}$ can be used to obtain matrix $\mathbf{B}$.

5.2 Sequential Linear Receiver

In this section, we propose a sequential type linear receiver as an alternative to the linear receiver characterized by matrix $\mathbf{B}$. The reason is that for ZF receiver in Chapter 3, in which case $\mathbf{B} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$, the number of multiplications to obtain $\mathbf{B}$ grows with N^3 due to inversion of $\mathbf{H}^H\mathbf{H}$. This makes the ZF receiver for oversampled uplink massive MIMO computationally prohibitive when the block length in a data packet goes high. One can propose using MRC receiver instead of ZF receiver, but it has been shown in [57] that MRC receiver suffers significantly from an error floor and performs worse in terms of error rate when oversampling is performed compared to the SR sampling case. Therefore, we seek to construct a sequential receiver that will provide the advantages that come with oversampling as in the linear ZF filter case and has a complexity that grows linearly with the block length N.

To derive a sequential linear receiver, we make some modifications in the signal model that we present in Section 5.1 such that the derived receiver does not wait for the whole quantized observation vector $\mathbf{r}$, which is of size βMN, to be obtained to update the estimate for the transmitted data symbol vector $\mathbf{x}$. In (5.12), the quantized receive vector $\mathbf{r}$ is expressed as a function of the data symbol vector $\mathbf{x}$. At this point, we define the observation vector at the time instant n, for which only the first n elements of the quantized receive vector $\mathbf{r}$ are observed, while the remaining $\beta MN - n$ elements are not observed yet. We denote this vector by $\mathbf{r}[n]$. In this case, the observation model can be expressed as

$$\mathbf{r}[n] = \mathcal{Q}(\mathbf{H}[n]\mathbf{x} + \mathbf{w}[n]), \tag{5.14}$$

where

$$\mathbf{r}[n] = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}^H, \tag{5.15}$$

$$\mathbf{w}[n] = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}^H, \tag{5.16}$$

$$\mathbf{H}[n] = \begin{bmatrix} \mathbf{h}[1]^H \\ \mathbf{h}[2]^H \\ \vdots \\ \mathbf{h}[n]^H \end{bmatrix}. \tag{5.17}$$

In (5.17), $\mathbf{h}[n]^H$ represents the nth row of matrix $\mathbf{H}$, whereas in (5.15) and (5.16), r_i and w_i are the ith element of the vectors $\mathbf{r}$ and $\mathbf{w}$, respectively. Denoting the soft estimate of vector $\mathbf{x}$ at the time instant $n-1$ as $\mathbf{x}'[n-1]$, which is calculated based on the observation vector at the time instant $n-1$, namely $\mathbf{r}[n-1]$, the aim is to update the estimate $\mathbf{x}'[n-1]$ to $\mathbf{x}'[n]$ using r_n. Such an estimator can be defined with the following update equations performed at time instant n.

$$\mathbf{k}[n] = \frac{\mathbf{M}[n-1]\hat{\mathbf{h}}[n]}{\sigma_n^2 + \hat{\mathbf{h}}[n]^H \mathbf{M}[n-1]\hat{\mathbf{h}}[n]}, \tag{5.18}$$

$$\mathbf{x}'[n] = \mathbf{x}'[n-1] + \mathbf{k}[n]\left[r_n - \hat{\mathbf{h}}[n]^H \mathbf{x}'[n-1]\right], \tag{5.19}$$

$$\mathbf{M}[n] = [\mathbf{I} - \mathbf{k}[n]\hat{\mathbf{h}}[n]^H]\mathbf{M}[n-1], \tag{5.20}$$

where $\mathbf{k}[n]$ can be denoted as the Kalman gain vector, $\mathbf{M}[n]$ as the MMSE matrix and $\hat{\mathbf{h}}[n]^H$ is the nth row of matrix $\hat{\mathbf{H}}$, which represents the estimated version of matrix $\mathbf{H}$. The execution of the update equations in (5.18)-(5.20) will also be referred to as the "n^{th} iteration step of the sequential receiver". The size of the Kalman gain vector $\mathbf{k}[n]$ is $NK \times 1$, whereas the size of the MMSE matrix $\mathbf{M}[n]$ is $NK \times NK$. When no data is observed yet, the symbol estimate $\hat{\mathbf{x}}[n]$ and the MMSE matrix $\mathbf{M}[n]$ should be initialized using the prior information of the symbol estimates as follows.

$$\mathbf{x}'[0] = \mathbf{E}[\mathbf{x}] = \mathbf{0}, \tag{5.21}$$

$$\mathbf{M}[0] = \mathbf{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}. \tag{5.22}$$

The equations (5.18)-(5.20) are the update equations for the sequential LMMSE for the unquantized observation model in (5.6) [61]. We propose this estimation technique to be directly used with the quantized observations. With such a sequential estimation scheme, the matrix inversion in the ZF receiver is avoided.

An important advantage of the proposed sequential estimator is its ability to provide estimates for the data vector without having to wait for all observations taken for the

whole data processing block (total number of observations for data processing block is βMN). For example, let the vector for the hard symbol estimates for the symbols transmitted at p^{th} symbol interval be denoted as $\hat{\mathbf{x}}^p$. It can be expressed as

$$\hat{\mathbf{x}}^p = \left[\begin{array}{cccc} \hat{x}_{(p-1)K+1} & \hat{x}_{(p-1)K+2} & \cdots & \hat{x}_{pK} \end{array} \right]^T, \tag{5.23}$$

where $\hat{x}_i$ corresponds to the ith element of the hard symbol estimate vector $\hat{\mathbf{x}}$ for the transmitted symbol vector $\mathbf{x}$ defined in (5.4). Let the soft estimates for the symbols transmitted in the pth symbol interval at the nth iteration step of the sequential estimation algorithm be denoted as $\mathbf{x}^{'}[n,p]$. It can be written as

$$\mathbf{x}'[n,p] = \left[x'_{(p-1)K+1}[n] \;\; x'_{(p-1)K+2}[n] \;\; \cdots \;\; x'_{pK}[n] \right]^T, \tag{5.24}$$

where $x'_i[n]$ corresponds to the ith element of the vector $\mathbf{x}'[n]$. Assuming that employed pulse shape decays to insignificant levels after L_p symbol durations, the hard symbol estimate vector for the symbols transmitted at p^{th} symbol interval $\hat{\mathbf{x}}^p$ can be found by mapping the elements in the soft estimate vector for the transmitted symbols at p^{th} symbol interval at the $M\beta(p+L_p)^{th}$ iteration step, namely $\mathbf{x}'[M\beta(p+L_p),p]$, to the minimum distance constellation point. In short, we make the hard decisions for the symbols transmitted at the nth symbol interval, whenever the observations for the $(n+L_p)^{th}$ symbol interval is taken. With this scheme, the maximum delay for the symbol decisions is L_pT for all transmitted symbols, while the delay for ZF receiver can be up to NT, which is in general significantly larger than L_pT.

Although the proposed sequential receiver characterized by (5.18)-(5.20) has the mentioned advantages, its complexity is still not low. For example, in (5.18), the complexity of the multiplication of matrix $\mathbf{M}[n-1]$, whose size is $NK \times NK$, with $\hat{\mathbf{h}}[n]$, which is a vector of size $NK \times 1$, grows with N^2. Repeating this multiplication for all iteration steps means that the complexity of the multiplications to estimate vector $\mathbf{x}$ at the end of all iterations grows with N^3, since the total number of iterations is βMN. Therefore, with the presented sequential receiver, the total number of multiplications still grows with N^3 similar to the ZF filter case. To reduce the complexity of the sequential receiver, we exploit the fact that the pulse shape decays to insignificant levels after a certain number of symbol durations (L_p), thus it is not necessary to update all symbol estimates for every observation. The low complexity version of the sequential receiver is characterized by the following update equations that are performed at each

iteration step.

$$\mathbf{k}_\ell[n] = \frac{\mathbf{M}_\ell[n-1]\hat{\mathbf{h}}_\ell[n]}{\sigma_n^2 + \hat{\mathbf{h}}_\ell[n]^H \mathbf{M}_\ell[n-1]\hat{\mathbf{h}}_\ell[n]}, \tag{5.25}$$

$$\mathbf{x}'_\ell[n] = \mathbf{x}'_\ell[n-1] + \mathbf{k}_\ell[n]\left[r_n - \hat{\mathbf{h}}_\ell[n]^H \mathbf{x}'_\ell[n-1]\right], \tag{5.26}$$

$$\mathbf{M}_\ell[n] = [\mathbf{I} - \mathbf{k}_\ell[n]\hat{\mathbf{h}}_\ell[n]^H]\mathbf{M}_\ell[n-1], \tag{5.27}$$

where

$$\mathbf{x}'_\ell[n] = \left[x'_{(z_n-L_p)K}[n] \;\; x'_{(z_n-L_p)K+1}[n] \;\; \cdots x'_{(z_n+L_p)K}[n]\right]^T, \tag{5.28}$$

$$\hat{\mathbf{h}}_\ell[n] = \left[\hat{h}_{(z_n-L_p)K}[n] \;\; \hat{h}_{(z_n-L_p)K+1}[n] \;\; \cdots \hat{h}_{(z_n+L_p)K}[n]\right]^T. \tag{5.29}$$

In (5.28) and (5.29), $x'_i[n]$ and $\hat{h}_i[n]$ corresponds to the ith element of the vectors $\mathbf{x}'[n]$ and $\hat{\mathbf{h}}[n]$, respectively. Moreover, the index z_n specifies the current symbol interval that the observations are being taken from at the nth iteration of the sequential receiver and is equal to $\lfloor (n-1)/M/\beta \rfloor + 1$, where $\lfloor . \rfloor$ is the floor function that gives the largest integer less than its operand. In this setting for the sequential receiver, the size of the Kalman gain $\mathbf{k}_\ell[n]$ and the MMSE matrix $\mathbf{M}_\ell[n]$ becomes $(2L_pK+1)\times 1$ and $(2L_pK+1)\times(2L_pK+1)$ and they can be initialized as in (5.21) and (5.22). For the low complexity version of the sequential receiver characterized by (5.25)-(5.27), the number of complex multiplications in the update equations (5.25)-(5.27) does not change with the block length N since the sizes of $\mathbf{h}_\ell[n-1]$, $\mathbf{x}'_\ell[n-1]$ and $\mathbf{M}_\ell[n-1]$ are $(2L_pK+1)\times 1$, $(2L_pK+1)\times 1$ and $(2L_pK+1)\times(2L_pK+1)$, respectively, which are all independent of N. Since there are βMN iterations, the number of multiplications grows with N compared to N^3 for the ZF receiver and the high complexity version of the sequential receiver characterized by (5.18)-(5.20).

5.3 Simulation Results

Number of users (K) and receive antennas (M) are taken to be 20 and 400, respectively. The block length (N) is selected to be 30 symbols. The roll-off factor (ρ) for the RRC pulse shape is taken to be 0.22. The parameter L_p is selected to be 4. The error rate performances of ZF receiver, whose complexity grows with N^3, and the low complexity sequential receiver characterized by its update equations in (5.25)-(5.27)

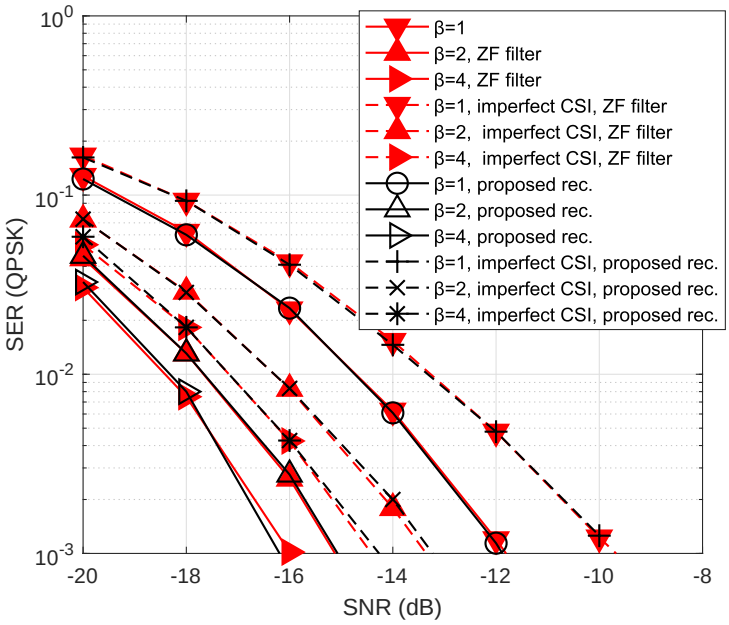


Figure 5.1: Simulation based SER vs. SNR curves for $M = 400$, $K = 20$, oversampling rate $\beta = 1, 2, 4$ with ZF and the proposed sequential receivers for perfect and imperfect CSI ($\sigma_h^2 = 0.2$).

are obtained by simulations and plotted in Fig. 5.1 for perfect and imperfect CSI cases when oversampling rates β is from 1 (no oversampling) to 4 (4 times oversampling). The SNR values on the plots are equal to $1/\sigma_n^2$.

The solid curves in Fig. 5.1 and Fig. 5.2 correspond to the curves obtained under perfect CSI ($\sigma_h^2 = 0$), whereas the dashed curves are for imperfect CSI case, for which $\sigma_h^2 = 0.2$. The black curves correspond to the performance of the high complexity ZF receiver, while the red curves represent the error-rate performance the proposed low complexity sequential receiver for both figures. As can be noted in Fig. 5.1, oversampling with ZF receiver provides up to 4 dB SNR advantage compared to the SR sampled case for both perfect and imperfect CSI cases when the SNR values to maintain a SER of 10^{-3} are considered, as pointed out in the Chapter 3. More importantly, the error rate performance of the proposed low complexity sequential receiver in this chapter is similar to the performance of the complex ZF receiver under both perfect and imperfect CSI cases. This means that the 4 dB SNR advantage provided by oversampling with ZF receiver compared to the SR sampled case is maintained with the proposed low complexity sequential receiver.

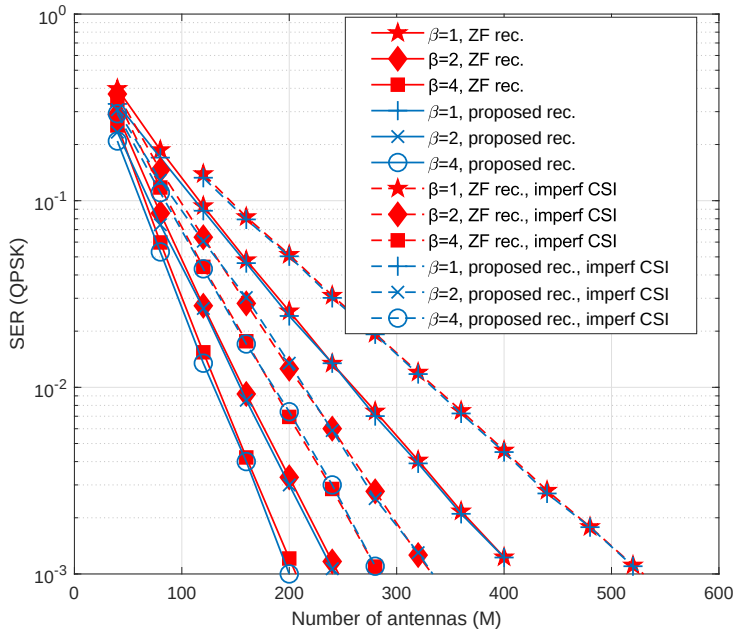


Figure 5.2: Simulation based SER vs number of receive antennas (M) for oversampling rate $\beta = 1, 2, 4$ with ZF and proposed sequential receivers when SNR=-12dB for perfect and imperfect CSI ($\sigma_h^2 = 0.2$).

We also present SER versus the number of receive antennas when SNR is fixed as -12 dB in Fig. 5.2. As can be inferred from Fig. 5.2, number of antennas necessary to maintain a SER of 10^{-3} can be halved by temporal oversampling with ZF receiver. This means that by oversampling, the form factor, power consumption and the overall cost of the MIMO array can be reduced significantly. Moreover, the pronounced advantages regarding the necessary number of antennas also prevail with the proposed low complexity receiver as its performance can be observed to be close to the ZF receiver in Fig. 5.2 for both perfect and imperfect CSI cases.

5.4 Conclusion

It has been shown in existing studies that temporal oversampling in 1-bit quantized uplink massive MIMO systems can provide significant advantages in terms of error rate performance and the necessary number of receive antennas required to maintain a certain error rate. However, the pronounced advantages of temporal oversampling are observed with a high complexity ZF receiver, whose complexity grows with N^3,

N being the block length, which is not implementable for long block lengths. In this chapter, a low complexity sequential receiver for temporally oversampled scenario is proposed whose complexity increases linearly with the block length. It has been observed that error rate performance of the proposed receiver is close to the performance of the complex ZF receiver for both perfect and imperfect CSI cases, thus it retains the advantages of temporal oversampling with a feasible receiver complexity. Moreover, the proposed receiver has shorter delay in providing the transmitted data symbol estimates.

CHAPTER 6

PERFORMANCE ANALYSIS OF QUANTIZED UPLINK MASSIVE MIMO-OFDM WITH OVERSAMPLING UNDER ADJACENT CHANNEL INTERFERENCE

6.1 Motivation and Contributions

In previous chapters, the benefits of temporal oversampling is presented for one-bit quantized massive MIMO systems. However, the question whether these advantages will also exist for the case when there is a strong interferer from an adjacent transmission band should still be answered. In this chapter, we make the performance analysis of quantized massive MIMO-OFDM structures when there is a strong interferer from an adjacent band.

The performance of heavily quantized massive MIMO with an interferer in an adjacent channel has not been examined in the related literature. However, such interference can be at significant levels due to near/far effect in a communication system in which users in the adjacent frequency band may be much closer to the receiver than the users in the desired band, thus, their signal may not be adequately suppressed by the receivers intending to extract the signals in the desired band. In fact, having the dynamic range to mitigate such interference is a key reason for using high-resolution ADCs in current systems [15]. Since distortion is large with low-resolution ADCs in practical use, there is a risk that such systems are practically nonoperational.

The study in this chapter is the first to analyze heavily quantized and OFDM modulated massive MIMO systems for an adjacent channel interference (ACI) scenario under frequency selective fading and channel estimation errors. It is also the first to analyze the performance of oversampling ADCs for such a scenario.

Other than the investigated scenario being different from the existing studies in the literature, the difference of this work compared to the aforementioned studies dealing with the analysis of quantized uplink massive MIMO systems with low-resolution ADCs are as follows. In [16, 17, 57], temporal oversampling and corresponding performance analysis is performed for an uplink massive MIMO system with one-bit ADCs in a single-carrier environment and flat fading, which results in a high complexity receiver in terms of baseband signal processing, whereas our study considers an OFDM system under frequency selective channel with low-resolution ADCs (not only one-bit), whose receiver complexity is changing almost linearly with oversampling. Another study [23] analyzes massive MIMO structures with one-bit ADCs under frequency selective fading. In that work, quantization noise is regarded as an uncorrelated distortion in time and space, which fails to hold when oversampling is performed or when the number of users or noise variance is not high [26, 30]. However, it will be seen that our analysis takes into account the temporal and spatial correlation in the quantization distortion, which enables an accurate analysis. Moreover, [30] provides an analysis for flat fading channels for massive MIMO structures with one-bit ADCs and provide a short section for the analysis of frequency selective channel case claiming that extension from flat fading channel case is straightforward. However, the sizes of the covariance matrices found for the quantized received signal for frequency selective case are $MN \times MN$, M and N being the number of antennas and block length (the number of samples in a coherence interval or in pilot duration). This large size makes their use in the performance analysis and channel estimation (covariance matrix inverse is used in channel estimation) infeasible in terms of computational complexity (even their storage in memory during simulations is problematic) unless the block length or number of antennas is very small, which is not the case for massive MIMO. There is a similar complexity problem in [26], where the matrix sizes involved in the calculation of the performance metrics are as large as $MN \times MN$. This problem is addressed in [25] by combining frequency domain operation with time domain operation for the calculation of necessary covariance matrices. However, many important points regarding the construction of the signal model using Bussgang decomposition or proofs regarding how to find correlation matrices in frequency domain from time domain matrices or vica versa are omitted. Furthermore, the calculation of a quantization noise covariance matrix in [25, Eqn.28] is valid only

when the quantization noise is uncorrelated over the time dimension, which does not hold for very low-resolution ADCs. Although the quantization noise covariance matrix calculation of [25] is shown to provide accurate results in [25], this is mostly due to the fact that the investigated ADC bit resolution in [25] (6 bits) is rather high. The more general version taking into account the time domain correlation is provided in Proposition 3 in this work. Moreover, the effect of system parameters such as the number of receive antennas or oversampling rate cannot be deduced from the signal-to-interference-noise-and-distortion ratio (SINDR) expressions in [25, 26], whereas we provide some approximate expressions for SINDR in Proposition 5, in which the effect of system parameters can easily be followed. Moreover, the analysis in [25, 26] is only performed for the perfect CSI case, whereas we propose a channel estimation algorithm and include the effect of imperfect CSI in this study. In summary, the contribution items are as provided below.

- This study is the first to analyze the performance of uplink massive MIMO systems with low-resolution ADCs in terms of error-rate and ergodic capacity under an ACI scenario. The analysis covers the frequency selective fading channel conditions. We obtain two types of analytical expressions for SINDR, one being more precise, whose accuracy is verified with simulations, and the other being less accurate but able to provide clear insights into the system performance and parameters. We show both analytically and with simulations that it is possible to combat ACI by increasing the number of receive antennas.

- We analyze the effect of oversampling in such systems and show analytically that oversampling is also effective to suppress ACI. We also show that significant performance gains can be obtained by oversampling either with simulations or theoretical analysis.

- We propose an LMMSE based channel estimation algorithm taking into account the effect of an adjacent channel interferer. The provided analysis is able to incorporate the effect of imperfect CSI on the system performance.

- We extend our analysis to multi-bit ADCs and discuss whether to increase the ADC resolution or oversampling rate by making comparisons in terms of error-rate performance while ADC power consumptions are kept constant.

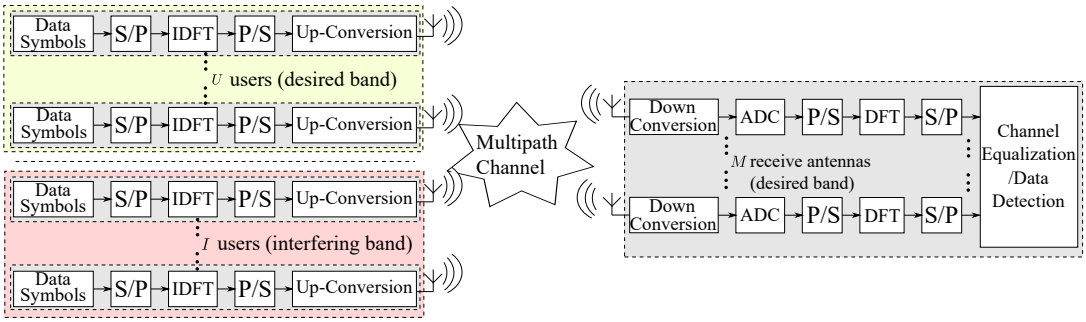


Figure 6.1: Multi-user uplink massive MIMO-OFDM block diagram in an interfering band scenario.

- The analysis in the work is general in that it can also be applied to the scenario where no ACI is present. For no ACI case, the analysis in the work
  - requires much less memory resources than the ones in [26,30] for SINDR calculations while providing closed-form expressions for quantization noise covariance matrix for one-bit ADCs as in (6.18) and (6.19) (more details for this item are mentioned previously),
  - takes into account the temporal and spatial correlation for the quantization noise, which are neglected in [23, 25], among which [23] does not cover the effect of oversampling,
  - can clearly show the impact of system parameters (such as number of antennas, oversampling rate, etc.) and imperfect CSI on the system performance unlike [25,26], where no channel estimation technique is proposed.

6.2 System Model

We consider the uplink scenario depicted in Fig. 6.1. It is assumed that K users send their information to a base station in an OFDM massive MIMO setting through a set of subcarriers that are assigned to them. This set of subcarriers will be denoted by $\mathcal{U}_D$ and will be referred to as the *desired band* in this script. The desired band users are illustrated with green background in Fig. 6.1. The receiver side in Fig. 6.1 is a typical OFDM receiver, in which the ADC block is assumed to have low-resolution in this study. Another group of I users, whose assigned set of subcarriers is denoted by the

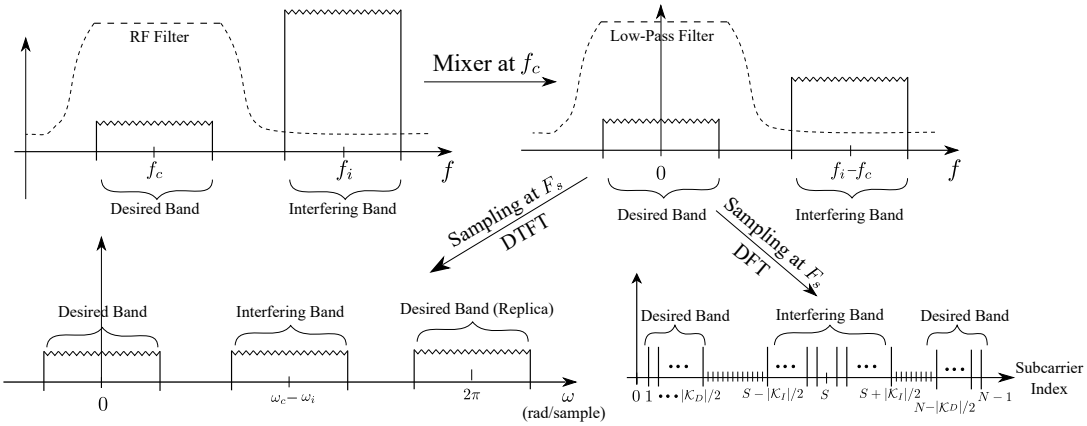


Figure 6.2: Example plots for the spectrum of the signals at various receiver stages.

set $\mathcal{U}_I$, is acting as an interfering source to the users in the desired band. These users are shaded with red background in Fig. 6.1. The set of subcarriers in $\mathcal{U}_I$ will also be referred to as *interfering band* in the remainder of this study. It should be noted that although the interference is from an adjacent band, it may not be suppressed enough due to near/far effect despite all analog filters involved in the down-conversion stages. Example plots for the spectrum of the signals at various points of a zero intermediate frequency (IF) receiver are provided in Fig. 6.2, where f_c and f_i are the carrier frequencies for the signals at the desired and interfering bands, respectively. The top left and right plots in Fig. 6.2 represent the power spectral densities (PSD) of the signals at the radio-frequency (RF) front end (just before the bandpass RF filter centered at f_c, for which the interfering band signal is much stronger compared to the desired band signal) and at the mixer output (before the low-pass filter (LPF) in the down-converter), respectively. Such a scenario can be considered as a typical long-term evolution (LTE) case in which different users are assigned to rectangular areas of resource blocks or subbands [62]. The discrete-time Fourier transform (DTFT) and discrete Fourier transform (DFT) of the sampled (but unquantized) signal are also shown at the bottom two plots in Fig. 6.2, where $\omega_c = 2\pi f_c/F_s$, $\omega_i = 2\pi f_i/F_s$ and $S = N(f_i - f_c)/(2F_s)$, F_s being the sampling rate. Moreover, it is also assumed that the receiver operates at a sampling rate fast enough to cover both the desired and interfering band to avoid any interference due to aliasing from the interfering band to the desired band. This is to study the isolated spectral leakage effect from the interfering to desired band owing to the non-linearity due to low-resolution ADCs, as ACI due to aliasing will occur even when there is no non-linearity, which is not the

focus of this work. The sets of users in the desired and interfering band are denoted by $\mathcal{K}_D = \{1, 2, \ldots, K\}$ and $\mathcal{K}_I = \{K+1, K+2, \ldots, K+I\}$, respectively.

6.3 Signal Model

We denote the complex data symbol of a user k transmitted at the uth subcarrier by $\tilde{s}_k[u]$, $u = 0, 1, \ldots, N-1$, where N is the DFT size. Not all subcarriers are occupied, that is, $\tilde{s}_k[u] = 0$ for $u \notin \mathcal{U}_D$ when $k \in \mathcal{K}_D$ or $u \notin \mathcal{U}_I$ when $k \in \mathcal{K}_I$, as shown in the bottom right of Fig. 6.2. The oversampling rate for the users in desired and interfering bands, β_D and β_I, are defined as a ratio of the total number of subcarriers N to the number of occupied subcarriers, that is, $\beta_D \triangleq N/|\mathcal{U}_D|$ and $\beta_I \triangleq N/|\mathcal{U}_I|$. Increasing N while $|\mathcal{U}_D|$ or $|\mathcal{U}_I|$ is fixed is termed as "oversampling" since we consider the case that the OFDM symbol duration NT_s is fixed, where T_s is the sampling period, requiring that the sampling rate $(1/T_s)$ is increased while N is increased. This also ensures that the transmission bandwidth of the desired channel users is kept the same when the oversampling rate β_D is increased, as the subcarrier spacing $1/(NT_s)$ and the number of occupied subcarriers $|\mathcal{U}_D|$ are fixed, which results in a fixed transmission bandwidth of $|\mathcal{U}_D|/(NT_s)$.

Following those definitions, the discrete-time signal of the kth user at the inverse DFT (IDFT) output, $\tilde{x}_{n,k}$, can be expressed by using (2.5) in Chapter 2 as

$$\tilde{x}_{n,k} = \begin{cases} \dfrac{\rho_d}{\sqrt{N}} \sum_{u \in \mathcal{U}_D} x_{u,k} e^{j2\pi(n-1)u/N} & \text{if } k \in \mathcal{K}_D, \\ \dfrac{\rho_i}{\sqrt{N}} \sum_{u \in \mathcal{U}_I} x_{u,k} e^{j2\pi(n-1)u/N} & \text{if } k \in \mathcal{K}_I, \end{cases} \tag{6.1}$$

for $n = 1, \ldots, N$. Here ρ_d and ρ_i are the average transmit power parameters for the desired or interfering band users. Moreover, data symbols have unit energy, that is, $\mathbf{E}[|\tilde{s}_k[u]|^2] = 1$. For simple equalization at the receiver side, a CP of length L_{cp} is added to the beginning of the OFDM symbol such that $\tilde{x}_{n,k} = \tilde{x}_{N+n,k}$ for $n = -L_{cp}+1, \ldots, 0$. It is required that $L_{cp} \geq L-1$, L being the number of channel taps[1].

The received signal at the mth antenna can be written by replacing T by T_s and K by

[1] L will be increased when the sampling rate $(1/T_s)$ is increased as the delay spread of the channels does not change with the sampling rate.

$K + I$ in (2.2) in Chapter 2 as

$$r_m(t) = \sum_{\ell=0}^{L-1} \sum_{k=1}^{K} \sum_{n=1}^{N} h_{m,k}[\ell] \tilde{x}_{n,k} p_c(t - (n-1)T_s - \ell T_s) + w_m(t), \tag{6.2}$$

where $p_c(t) = sinc(t/T_s)$. It is assumed that the sampling rate of the receiver is the same as that of the transmitter. As $sinc(t/T_s) = 0$ for $t = nT_s$ for $n \neq 0$, the discrete-time received signal at the mth antenna, namely $r_m[n] = r_m(nT_s)$, can be expressed as follows:

$$r_m[n] = r_m(nT_s) = \sum_{k=1}^{K+I} \sum_{\ell=0}^{L-1} h_{m,k}[\ell] s_k[n-\ell] + w_m[n], \tag{6.3}$$

where $s_k[n] = \tilde{x}_{n+1,k}$, $n = 0, \ldots, N-1$, $w_m[n] = w_m(nT_s)$. We assume that the channel coefficients $h_{m,k}[\ell]$ have complex Gaussian distribution and are uncorrelated, that is, $\mathbf{E}[h_{m_1,k_1}[\ell_1] h_{m_2,k_2}[\ell_2]^*] = p_{k_1}[\ell_1]\delta[\ell_1 - \ell_2]\delta[k_1 - k_2]\delta[m_1 - m_2]$, where $p_k[\ell]$ is the power delay profile of the channel between user k and the receive antennas satisfying $\sum_{\ell=0}^{L-1} p_k[\ell] = 1 \, \forall k$. The justification behind this assumption is discussed in Chapter 2. In addition, the noise samples $w_m[n]$ are also assumed to be uncorrelated. A more compact version of (6.3) is

$$\mathbf{r}[n] = \sum_{\ell=0}^{L-1} \mathbf{H}[\ell] \mathbf{s}[n-\ell] + \mathbf{w}[n], \tag{6.4}$$

where $\mathbf{H}[\ell]$ is an $M \times (K + I)$ matrix whose element at the mth row and the kth column is $h_{m,k}[\ell]$. Moreover, $\mathbf{s}[n]$ is a column vector whose k^{th} element is $s_k[n]$. Furthermore, $\mathbf{w}[n]$ and $\mathbf{r}[n]$ are column vectors whose m^{th} element is equal to $w_m[n]$ and $r_m[n]$, respectively.

The signal at the output of the one-bit quantizer (the case for the higher bit resolutions is presented in Section 6.5), namely $\mathbf{d}[n]$, can be written in terms of its input, $\mathbf{r}[n]$ as

$$\mathbf{d}[n] = \mathrm{sign}(\mathrm{Re}(\mathbf{r}[n])) + j\mathrm{sign}(\mathrm{Im}(\mathbf{r}[n])), \tag{6.5}$$

where $\mathrm{sign}(.)$ is the signum function. The DFT of the quantizer output is also taken to obtain the input signal to the channel equalization and data detection block, namely $\tilde{\mathbf{d}}[u]$, as follows:

$$\tilde{\mathbf{d}}[u] = \sum_{n=0}^{N-1} \mathbf{d}[n] e^{-j2\pi nu/N}. \tag{6.6}$$

How to obtain the data estimates based on $\tilde{\mathbf{d}}[u]$ will be considered in the next section.

6.4 Performance Analysis

For a tractable analysis of the non-linear system with one-bit ADCs, the Bussgang decomposition [30], which enables a linear input-output relation for a non-linear system, will be employed. Before that, it is necessary to reexpress (6.4) as

$$\underline{\mathbf{r}} = \underline{\mathbf{H}}\,\underline{\mathbf{s}} + \underline{\mathbf{w}}, \tag{6.7}$$

$\underline{\mathbf{r}} = \left[\mathbf{r}\,[N-1]^T\ \mathbf{r}\,[N-2]^T\ \cdots\ \mathbf{r}\,[0]^T\right]^T$, $\underline{\mathbf{s}} = \left[\mathbf{s}\,[N-1]^T\ \mathbf{s}\,[N-2]^T\ \cdots\ \mathbf{s}\,[0]^T\right]^T$, $\underline{\mathbf{w}} = \left[\mathbf{w}\,[N-1]^T\ \mathbf{w}\,[N-2]^T\ \cdots\ \mathbf{w}\,[0]^T\right]^T$, and $\underline{\mathbf{H}}$ is a block circulant matrix of size $NM \times N(K+I)$ that can be expressed as follows:

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{H}[0] & \mathbf{H}[1] & \cdots & \mathbf{H}[L-1] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}[0] & \mathbf{H}[1] & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}[0] & \cdots & \cdots & \mathbf{H}[L-1] \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \cdots \\ \mathbf{H}[2] & \mathbf{H}[1] & \cdots & \mathbf{H}[L-2] & \cdots & \mathbf{H}[0] & \mathbf{H}[1] \\ \mathbf{H}[1] & \mathbf{H}[2] & \cdots & \mathbf{H}[L-1] & \cdots & 0 & \mathbf{H}[0] \end{bmatrix} \tag{6.8}$$

This is the MIMO extension of the circulant channel matrix defined for a single-input single-output (SISO) OFDM scenario in [3]. According to the Bussgang decomposition [30, 63],

$$\underline{\mathbf{d}} = \underline{\mathbf{A}}\,\underline{\mathbf{r}} + \underline{\mathbf{q}}, \tag{6.9}$$

where $\underline{\mathbf{A}} = \mathbf{C}^H_{\underline{\mathbf{d}}\,\underline{\mathbf{r}}}\mathbf{C}^{-1}_{\underline{\mathbf{r}}}$, in which $\mathbf{C}_{\underline{\mathbf{r}}} = \mathbf{E}[\underline{\mathbf{r}}\,\underline{\mathbf{r}}^H]$ and $\mathbf{C}_{\underline{\mathbf{d}}\,\underline{\mathbf{r}}^H} = \mathbf{E}[\underline{\mathbf{d}}\,\underline{\mathbf{r}}^H]$, Moreover, $\underline{\mathbf{q}}$ is the equivalent quantization noise vector, $\underline{\mathbf{d}} = \left[\mathbf{d}\,[N-1]^T\ \mathbf{d}\,[N-2]^T\ \cdots\ \mathbf{d}\,[0]^T\right]^T$. The size of $\underline{\mathbf{A}}$ is $NM \times NM$. This choice of $\underline{\mathbf{A}}$ minimizes the variance of the quantizer noise or equivalently makes $\underline{\mathbf{r}}$ uncorrelated with $\underline{\mathbf{q}}$. For a one-bit quantizer, assuming zero-mean Gaussian inputs[2], the following holds [30]:

$$\underline{\mathbf{A}} = \sqrt{\frac{4}{\pi}}\mathrm{diag}(\mathbf{C}_{\underline{\mathbf{r}}})^{-0.5} = \sqrt{\frac{4}{\pi}}\mathrm{diag}\left(\underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{s}}}\underline{\mathbf{H}}^H + N_o\mathbf{I}\right)^{-0.5}, \tag{6.10}$$

[2] This assumption is approximately true even when the transmitted symbols are from a finite cardinality set rather than being Gaussian distributed. The ADC input at the mth antenna can be written as sum of $\mathcal{S}_K^m$ and $\mathcal{S}_I^m$, where $\mathcal{S}_K^m$ is the sum of $K|\mathcal{U}_\mathcal{D}|L$ independent identically distributed (i.i.d.) signals with finite variance from the desired band, and $\mathcal{S}_I^m$ is the sum of $I|\mathcal{U}_\mathcal{I}|L$ i.i.d. signals from the interfering band. Due to the central limit theorem, $\mathcal{S}_K^m$ and $\mathcal{S}_I^m$ converge to Gaussian as $K|\mathcal{U}_\mathcal{D}|L$ and $I|\mathcal{U}_\mathcal{I}|L$ grow large, so does the ADC input $\mathcal{S}_K^m + \mathcal{S}_I^m$. It can be shown by the Berry-Essen inequality that the difference between the CDFs of $\mathcal{S}_K^m$ or $\mathcal{S}_I^m$ and the CDF of a Gaussian random variable with the same mean and variance is always less than 0.02, even when K and I are as low as 4, $L = 3$ and $|\mathcal{U}_\mathcal{D}|$ or $|\mathcal{U}_\mathcal{I}|$ is 128. Since CDFs are between 0 and 1, an error of 0.02 is negligable. For higher $I|\mathcal{U}_\mathcal{I}|L$ or $K|\mathcal{U}_\mathcal{D}|L$, this error will be much less (error decreases with $\sqrt{K|\mathcal{U}_\mathcal{D}|L}$ or $\sqrt{I|\mathcal{U}_\mathcal{I}|L}$).

where $\mathbf{R}_{\underline{\mathbf{s}}} = \mathbf{E}[\underline{\mathbf{s}}\underline{\mathbf{s}}^H]$. Although a closed form expression as in (6.10) is available to calculate the matrix $\underline{\mathbf{A}}$, its calculation is not simple, whose complexity is in the order of N^3M^2, which is very large for a typical massive MIMO scenario. Therefore, an alternative low-complexity approach will be presented. Since $\underline{\mathbf{A}}$ is a diagonal matrix, (6.9) can be modified as

$$\mathbf{d}[n] = \mathbf{A}[n]\mathbf{r}[n] + \mathbf{q}[n], \tag{6.11}$$

for $n = 1, \ldots, N$, where $\mathbf{A}[n]$ is a diagonal matrix with diagonal elements equal to a set of diagonal elements of $\underline{\mathbf{A}}$, which are, $[(\underline{\mathbf{A}})_{n,n}\ (\underline{\mathbf{A}})_{n+1,n+1} \cdots\ (\underline{\mathbf{A}})_{n+M-1,n+M-1}]$. Moreover, as $\mathbf{r}[n]$ is stationary owing to the addition of the cyclic prefix and $\mathbf{A}[n]$ is diagonal, it can be shown that $\mathbf{A}[n] = \mathbf{A}[n'] \triangleq \mathbf{A}, \forall n, n'$, thus

$$\mathbf{d}[n] = \mathbf{A}\mathbf{r}[n] + \mathbf{q}[n], \tag{6.12}$$

$\mathbf{A} = \sqrt{4/\pi}\mathrm{diag}\left(\mathbf{C}_{\mathbf{r}}[0]\right)^{-0.5}$ is an $M \times M$ matrix, where $\mathbf{C}_{\mathbf{r}}[m] \triangleq \mathbb{E}\left[\mathbf{r}[n]\mathbf{r}[n-m]^H\right]$. It is still hard to find $\mathbf{C}_{\mathbf{r}}[0]$ in the time domain using (6.4) as there exists correlation between the time domain symbols $\mathbf{s}[n]$ due to oversampling. It will also be more convenient to work in the frequency domain as the final detection of the data symbols will be performed in that domain, thus any SINDR expression to be used in the analysis should be found for the frequency domain observations. Therefore, the analysis continues by taking the DFT of both sides of (6.12), yielding

$$\tilde{\mathbf{d}}[u] = \mathbf{A}\tilde{\mathbf{r}}[u] + \tilde{\mathbf{q}}[u], \tag{6.13}$$

where $\tilde{\mathbf{r}}[u]$ has a simple expression that can be found by using (6.4) and considering the circulant property of the channel convolution matrix due to the addition of the cyclic prefix as

$$\tilde{\mathbf{r}}[u] = \begin{cases} \rho_d\sqrt{N}\tilde{\mathbf{H}}[u]\tilde{\mathbf{s}}[u] + \tilde{\mathbf{w}}[u], & \text{if } u \in \mathcal{U}_D, \\ \rho_i\sqrt{N}\tilde{\mathbf{H}}[u]\tilde{\mathbf{s}}[u] + \tilde{\mathbf{w}}[u], & \text{if } u \in \mathcal{U}_I, \\ 0 & \text{otherwise}, \end{cases} \tag{6.14}$$

where $\tilde{\mathbf{s}}[u] \triangleq [\tilde{s}_1[u]\ \tilde{s}_2[u]\ \cdots\ \tilde{s}_{K+I}[u]]$, $u = 0, 1, \ldots, N-1$. Moreover, $\tilde{\mathbf{r}}[u] = \sum_{n=0}^{N-1} \mathbf{r}[n]e^{-j2\pi nu/N}$, $\tilde{\mathbf{H}}[u] = \sum_{n=0}^{N-1} \mathbf{H}[n]e^{-j2\pi nu/N}$, $\tilde{\mathbf{w}}[u] = \sum_{n=0}^{N-1} \mathbf{w}[n]e^{-j2\pi nu/N}$. Next, we define $\mathbf{C}_{\tilde{\mathbf{r}}[u]} \triangleq \mathbb{E}\left[\tilde{\mathbf{r}}[u]\tilde{\mathbf{r}}[u]^H\right]$. Since $\mathbf{E}[\tilde{\mathbf{r}}[u]\tilde{\mathbf{r}}[u']^H] = \mathbf{0}$ for $u \neq u'$, $\mathbf{C}_{\mathbf{r}}[m]$ can be found according to the following proposition.

Proposition 2: $\mathbf{C_r}[m]$ *can be computed from* $\mathbf{C}_{\tilde{\mathbf{r}}[u]}$ *as*

$$\mathbf{C_r}[m] = \frac{1}{N^2}\sum_{u=0}^{N-1}\mathbf{C}_{\tilde{\mathbf{r}}[u]}e^{j2\pi mu/N}. \tag{6.15}$$

Proof: See Appendix B.1. ■

$\mathbf{C_r}[m]$ can be calculated by taking the IDFT of $\mathbf{C}_{\tilde{\mathbf{r}}[u]}$, which can be found using (6.14) as

$$\mathbf{C}_{\tilde{\mathbf{r}}[u]} = \begin{cases} \rho_d^2 N\tilde{\mathbf{H}}[u]\tilde{\mathbf{H}}[u]^H + NN_o\mathbf{I}, & \text{if } u \in \mathcal{U}_D, \\ \rho_i^2 N\tilde{\mathbf{H}}[u]\tilde{\mathbf{H}}[u]^H + NN_o\mathbf{I}, & \text{if } u \in \mathcal{U}_I, \\ 0, & \text{otherwise.} \end{cases} \tag{6.16}$$

What remains is to find the covariance matrix of the quantization distortion $\tilde{\mathbf{q}}[u]$, namely $\mathbf{C}_{\tilde{\mathbf{q}}[u]} = \mathbf{E}[\tilde{\mathbf{q}}[u]\tilde{\mathbf{q}}[u]^H]$. Consider the quantization noise vector $\underline{\mathbf{q}}$ in (6.9). $\mathbf{C}_{\underline{\mathbf{q}}} = \mathbf{E}[\underline{\mathbf{q}}\,\underline{\mathbf{q}}^H]$ is given by

$$\mathbf{C}_{\underline{\mathbf{q}}} = \mathbf{C}_{\underline{\mathbf{d}}} - \underline{\mathbf{A}}\mathbf{C}_{\underline{\mathbf{r}}}\underline{\mathbf{A}}^H. \tag{6.17}$$

The matrix sizes in (6.17) are $MN \times MN$, which can be very large, and require vast amount of memory resources for the computations (for instance, with $M = 64, N = 1024$, each matrix in (6.17) requires about 32 GB space for double precision number format). Such large matrices are also present in the studies [26, 30]. Therefore, an alternative method will be proposed to work with matrices of feasible sizes. From the block Toeplitz structure of $\mathbf{C}_{\underline{\mathbf{q}}}$ and the fact that $\underline{\mathbf{A}}$ is a diagonal matrix, it can be shown that

$$\mathbf{C_q}[m] = \mathbf{C_d}[m] - \mathbf{A}\mathbf{C_r}[m]\mathbf{A}^H, \tag{6.18}$$

$$\begin{aligned}\mathbf{C_d}[m] = \frac{4}{\pi}\Big(&\operatorname{asin}\left(\mathbf{D_r}[m]^{-\frac{1}{2}}\operatorname{Re}(\mathbf{C_r}[m])\mathbf{D_r}[m]^{-\frac{1}{2}}\right) \\ &+ j\operatorname{asin}\left(\mathbf{D_r}[m]^{-\frac{1}{2}}\operatorname{Im}(\mathbf{C_r}[m])\mathbf{D_r}[m]^{-\frac{1}{2}}\right)\Big),\end{aligned} \tag{6.19}$$

where $\mathbf{D_r}[m] = \operatorname{diag}(\mathbf{C_r}[m])$ and $m = 0, 1, \ldots, N-1$. (6.19) is a result of the *arcsine law* [64]. Note that the matrix sizes in (6.18) and (6.19) are $M \times M$ and the memory requirement to hold the matrices in (6.18) is N times smaller compared to (6.17). Moreover, it can be noted that the temporal and spatial correlation of quantization noise is taken into account since $\mathbf{C_q}[m]$ can be calculated for $m \neq 0$ using

(6.18) and is not necessarily a diagonal matrix. Now that every matrix in (6.18) is known, $\mathbf{C}_{\mathbf{q}}[m]$ can be calculated. The next step is to find $\mathbf{C}_{\tilde{\mathbf{q}}[u]} = \mathbf{E}[\tilde{\mathbf{q}}[u]\tilde{\mathbf{q}}[u]^H]$ from $\mathbf{C}_{\mathbf{q}}[m]$, which is performed in the following proposition.

Proposition 3: $\mathbf{C}_{\mathbf{q}[u]}$ *can be computed from* $\mathbf{C}_{\mathbf{q}}[m]$ *as*

$$\mathbf{C}_{\mathbf{q}[u]} = \mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[m]\right\} + \mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[m]\right\}^* - N\mathbf{C}_{\tilde{\mathbf{q}}}[0], \tag{6.20}$$

where $\mathbf{\Gamma}[m] \triangleq (N-m)\mathbf{C}_{\mathbf{q}}[m]$ *and* $\mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[m]\right\} = \sum_{m=0}^{N-1}\mathbf{\Gamma}[m]e^{-j2\pi mu/N}$.

Proof: See Appendix B.2. ■

6.4.1 Data Detection

For data detection, ZF combining is applied to the DFT output $\tilde{\mathbf{d}}[u]$ to obtain data estimates $\hat{\mathbf{x}}[u]$ as

$$\hat{\mathbf{x}}[u] = \hat{\mathbf{B}}[u]\tilde{\mathbf{d}}[u], \tag{6.21}$$

for $u \in \mathcal{U}_D$. In (6.21), $\hat{\mathbf{B}}[u] = \left(\hat{\mathbf{H}}[u]^H\hat{\mathbf{H}}[u]\right)^{-1}\hat{\mathbf{H}}[u]^H$, where $\hat{\mathbf{H}}[u]$ is the estimate for $\tilde{\mathbf{H}}[u]$. The details about the channel estimation is provided in Section 6.4.2. Using (6.13), (6.14) and (6.21), the kth element of $\hat{\mathbf{x}}[u]$, namely $\hat{x}_k[u]$, can be found as

$$\hat{x}_k[u] = g_k[u] + i_k[u] + n_k[u] + q_k[u], \tag{6.22}$$

where

$$g_k[u] = g_k[u]'\tilde{s}_k[u],\ n_k[u] = \hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{w}}[u], q_k[u] = \hat{\mathbf{b}}_k[u]^H\mathbf{q}[u], \tag{6.23}$$

$$i_k[u] = \rho_d\sqrt{N}\left[\sum_{z\neq k, z\in\mathcal{K}_D}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_z[u]\tilde{s}_z[u] + \sum_{z\neq k, z\in\mathcal{K}_D}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{e}}_z[u]\tilde{s}_z[u]\right], \tag{6.24}$$

where $g_k[u]' = \rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{h}}_k[u]$. In (6.22), $g_k[u]$ corresponds to the signal part, whereas, $i_k[u]$ term in (6.22) contains the interference from other users (the left summation in the $i_k[u]$ expression) and the distortion caused by imperfect CSI (the right summation in the $i_k[u]$ expression). Furthermore, $n_k[u]$ and $q_k[u]$ correspond to the distortion caused by thermal noise and quantization, respectively. Moreover, $\hat{\mathbf{b}}_k[u]^H$ is the kth row of the ZF combiner $\hat{\mathbf{B}}[u]$. Furthermore, $\tilde{\mathbf{h}}_k[u]$ and $\hat{\mathbf{h}}_k[u]$ are equal to

the kth column of the matrices $\tilde{\mathbf{H}}[u]$ and $\hat{\mathbf{H}}[u]$, respectively. In addition, $\tilde{\mathbf{e}}_k[u]$ corresponds to the kth column of the channel error matrix $\tilde{\mathbf{H}}[u] - \hat{\mathbf{H}}[u]$.

A lower bound on the ergodic capacity per user, for which the receiver has access to the side information $\hat{\mathbf{H}} \triangleq \{\hat{\mathbf{H}}[0], \hat{\mathbf{H}}[1], \ldots, \hat{\mathbf{H}}[N-1]\}$, an SINDR expression for the data symbol of kth user at the uth subcarrier, namely $\gamma_k[u]$, which is defined in [2, eq. (2.46)], is found using (6.22)-(6.24) in the following proposition.

Proposition 4: A lower bound on the ergodic capacity per user, treating $\hat{\mathbf{H}}$ as side information, denoted by R, can be calculated as

$$R = \frac{1}{K|\mathcal{U}_D|}\mathbf{E}_{\hat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{u\in\mathcal{U}_D}\log_2\left(1+\gamma_k[u]\right)\right\}, \tag{6.25}$$

where

$$\begin{aligned}\gamma_k[u] &\triangleq \frac{|\mathbf{E}[g_k[u]'|\hat{\mathbf{H}}]|^2}{\mathit{Var}\left[g_k[u]'|\hat{\mathbf{H}}\right] + \mathit{Var}\left[i_k[u]+n_k[u]+q_k[u]|\hat{\mathbf{H}}\right]} \\ &= \frac{\rho_d^2 N|\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_k[u]|^2}{I_k[u]+N_k[u]+Q_k[u]},\end{aligned} \tag{6.26}$$

$$I_k[u] = \rho_d^2 N\sum_{z\neq k, z\in\mathcal{K}_D}\left[|\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_z[u]|^2 + K\sigma_e^2\rho_d^2||\hat{\mathbf{b}}_k[u]^H\mathbf{A}||^2\right], \tag{6.27}$$

$$N_k[u] = NN_o||\hat{\mathbf{b}}_k[u]^H\mathbf{A}||^2, Q_k[u] = \hat{\mathbf{b}}_k[u]^H\mathbf{C}_{\mathbf{q}[u]}\hat{\mathbf{b}}_k[u], \tag{6.28}$$

when $\hat{\mathbf{H}}$ is the LMMSE channel estimate. In (6.27), σ_e^2 is the LMMSE channel estimation error variance, which is as defined in Section 6.4.2.

Proof: See Appendix B.3. ■

Moreover, to find the bit error rate (BER) for gray coded P-ary phase shift keying modulation (PSK with modulation size P), we approximate $\hat{\mathbf{x}}[u]$ as a complex Gaussian random variable, so that [3]

$$\text{BER} \approx \frac{1}{K|\mathcal{U}_D|}\mathbf{E}_{\hat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{u\in\mathcal{U}_D}\frac{2}{\log_2(P)}\times\left(1-\Phi\left(\sqrt{\gamma_k[u]\log_2(P)}\sin\left(\frac{\pi}{P}\right)\right)\right)\right\}. \tag{6.29}$$

The BER calculation formula in (6.29) is applicable for a multi-band interference environment, as the calculation of the SINDR $\gamma_k[u]$ involves the received power from any interfering band. As detailed in Footnote 1 of Section 6.3, the approximation error in assuming Gaussian inputs for the quantizer is very limited, thus R and the BER expressions in (6.25) and (6.29) calculated using $\gamma_k[u]$ in (6.26) are expected to approximate the simulation based results precisely ($\gamma_k[u]$ is the precise SINDR expression mentioned in Section 6.1). Although (6.25) and (6.29) are useful to calculate the ergodic capacity and BER, they are not able to provide clear insights into the system performance and system parameters such as M, N or ρ_d^2/ρ_i^2. Therefore, we propose a more tractable approximation of $\gamma_k[u]$ in Proposition 5, in which the conditioning on $\hat{\mathbf{H}}$ will be dropped. From an information theoretic view, this will correspond to the case that the channel estimates are used to perform ZF combining by a first party, but the channel estimate knowledge is not conveyed to a second party, which performs error correction decoding on the ZF output without the knowledge of channel estimates [2]. The corresponding use-and-then-forget ergodic capacity bound, namely R', is found in the following proposition.

Proposition 5: Use-and-then-forget ergodic capacity bound R' can be computed as

$$R' = \frac{1}{K|\mathcal{U}_D|} \sum_{k=1}^{K} \sum_{u \in \mathcal{U}_D} log_2 \left(1 + \gamma_k'[u]\right), \tag{6.30}$$

$$\begin{aligned} \gamma_k'[u] &= \frac{|\mathbf{E}[g_k[u]']|^2}{Var[g_k[u]'] + Var[i_k[u] + n_k[u] + q_k[u]]} \\ &\approx \frac{\rho_d^2(1-\sigma_e^2)(M-K)G^2}{(K\sigma_e^2\rho_d^2G^2 + 2 - 4/\pi + N_oG^2)}, \end{aligned} \tag{6.31}$$

in which $G = 2/\sqrt{\pi}\left((|\mathcal{U}_D|K\rho_d^2 + |\mathcal{U}_I|I\rho_i^2)/N + N_o\right)^{-0.5}$. The approximation error goes to zero as L gets larger and $|\mathcal{U}_D| + |\mathcal{U}_I|$ approach to N (as oversampling rates gets lower) when $\rho_d^2 \approx \rho_i^2$.

Proof: See Appendix B.4. ■

Note that there is no need for any Monte-Carlo based simulation to obtain $\gamma_k'[u]$, since it can be calculated by just plugging the system parameters into the right-most expression in (6.31). $\gamma_k'[u]$ can be used in place of $\gamma_k[u]$ in (6.29) or in (6.30)

to calculate BER and achievable rate, again without any Monte-Carlo simulations (as averaging over $\hat{\mathbf{H}}$ is already performed by analysis to obtain $\gamma'_k[u]$). Owing to the simple form of $\gamma'_k[u]$ in (6.31), insights into the system performance and parameters can be obtained as follows. From Proposition 5, it is obvious that it is always possible to increase the SINDR by increasing the number of antennas M. The impact of the oversampling rate can be deduced by considering the case $|\mathcal{U}_D|$ and $|\mathcal{U}_I|$ are fixed as the block length N increases which by definition is equivalent to an increase in the oversampling rates β_I and β_D. It can be shown that $\gamma'_k[u]$ is increasing with N by considering $\gamma'_k[u] = \rho_d^2(M-K)(1-\sigma_e^2)\overline{\gamma}_k[u]$, where $\overline{\gamma}_k[u] = G^2/\left(2-4/\pi+(N_o+K\sigma_e^2\rho_d^2)\,G^2\right)$. Since G and G^2 are increasing with N, it follows that $\overline{\gamma}_k[u] = G^2/\left(2-4/\pi+(N_o+K\sigma_e^2\rho_d^2)\,G^2\right)$ also increases with N, which in turn means that $\gamma'_k[u]$ increases with N. The channel estimation error will be calculated in (6.43) in the next section, which can similarly be shown to decrease with the oversampling rates.

6.4.2 Channel Estimation

In this section, the details of channel estimation under quantization will be presented. There are many channel estimation techniques for massive MIMO systems with low-resolution ADCs [12, 23, 36, 53, 59]. In none of those studies, a channel estimation scheme under ACI is discussed. In this study, we will propose an LMMSE channel estimation based on Bussgang decomposition. The estimation technique is an extension of the channel estimation technique in [23]. For the channel estimation phase, orthogonal pilot sequences are transmitted at some of the subcarriers involved. The set of subcarriers for which pilot signals are transmitted is denoted by $\mathcal{U}_P$. Moreover, it will be assumed that the interferers at the adjacent channel band will be transmitting data symbols from a finite cardinality set through the subcarriers in $\mathcal{U}_I$ during the channel estimation phase of the users in the desired band. The transmitted time domain pilot signal of the kth user of length N_p, where $k \in \mathcal{K}_D$, can be expressed as follows:

$$p_k[n] = \begin{cases} \dfrac{\rho_p}{\sqrt{N_p}} \sum_{u\in\mathcal{U}_P} \tilde{\theta}_k[u] e^{j2\pi nu/N_p}, & \text{if } k \in \mathcal{K}_D, \\ \dfrac{\rho_i}{\sqrt{N_p}} \sum_{u\in\mathcal{U}_I} \tilde{s}_k[u] e^{j2\pi nu/N_p}, & \text{if } k \in \mathcal{K}_I, \end{cases} \tag{6.32}$$

$$\tilde{\theta}_k[u] = \begin{cases} 0, & \text{if } (u \bmod K) + 1 \neq k, \\ \sqrt{K} e^{j\phi_k[u]}, & \text{if } (u \bmod K) + 1 = k. \end{cases} \tag{6.33}$$

Here, the phases $e^{j\phi_k[u]}$ are known by the base station. The selection of these phases affect the estimation performance. They are selected from the uniform distribution as suggested in [23] since when they are selected as constant, that is, when $e^{j\phi_k[u]} = C \; \forall k, u$, the transmitted signal by k^{th} user $p_k[n] \neq 0$ only when $n = \mu N_p/, \mu \in \mathbb{Z}$. Otherwise, $p_k[n] = 0$. This means that users do not transmit anything most of the time, which limits the average transmit power due to the peak power limitation of the power amplifiers in the transmitter side. Introduction of non-constant phases, one example of which is when they are selected from uniform distribution $(0, 2\pi)$, avoids this problem. Since the channel convolution matrix is circulant due to CP, the received signal at the mth antenna and the uth subcarrier can be expressed as follows:

$$\tilde{y}_m[u] = \begin{cases} \rho_p \sqrt{N_p K} \sum_{k \in \mathcal{K}_D} \tilde{h}_{m,k}[u] \tilde{\theta}_k[u] + z_m[u], & \text{if } u \in \mathcal{U}_P, \\ \rho_i \sqrt{N_p K} \sum_{k \in \mathcal{K}_I} \tilde{h}_{m,k}[u] \tilde{s}_k[u] + z_m[u], & \text{if } u \in \mathcal{U}_I, \end{cases} \tag{6.34}$$

where $\tilde{s}_k[u]$'s for $u \in \mathcal{U}_I$, $k \in \mathcal{K}_I$ represent the random data symbols transmitted by the interfering band users, $\tilde{h}_{m,k}[u]$ is the element of matrix $\tilde{\mathbf{H}}[u]$ at its mth row and k^{th} column, and $z_m[n]$ represents the additive white noise term at the mth receive antenna of spectral density $N_p N_o$. Due to (6.33), it can be written that

$$\tilde{y}_m[u] = \begin{cases} \rho_p \sqrt{N_p K} \tilde{h}_{m,f(u)}[u] e^{j\phi_{f(u)}[u]} + z[u], & \text{if } u \in \mathcal{U}_P, \\ \rho_i \sqrt{N_p K} \sum_{k \in \mathcal{K}_I} \tilde{h}_{m,k}[u] \tilde{s}_k[u] + z[u], & \text{if } u \in \mathcal{U}_I, \end{cases} \tag{6.35}$$

where $f(u) = (u \bmod K) + 1$. Defining the quantized observation vector $v_m[n] \triangleq \text{sign}(\text{Re}\{y_m[n]\}) + j\text{sign}(\text{Im}\{y_m[n]\})$ in time domain, the quantized observation $\tilde{v}_m[u]$ in the frequency domain can be expressed as

$$\tilde{v}_m[u] = \sum_{u=0}^{N_p-1} v_m[n] e^{j2\pi nu/N_p}. \tag{6.36}$$

Here $y_m[n]$ is the IDFT of $\tilde{y}_m[u]$. Defining $\mathbf{v}[n] \triangleq [v_1[n] \; v_2[n] \; \cdots \; v_M[n] \;]^T$ and $\mathbf{y}[n] \triangleq [\; y_1[n] \; y_2[n] \; \cdots \; y_M[n] \;]^T$, it can be written that

$$\mathbf{v}[n] = \mathbf{A}'\mathbf{y}[n] + \mathbf{q}'[n], \tag{6.37}$$

where the selection $\mathbf{A}' = \sqrt{4/\pi}\text{diag}\left(\mathbf{C}_y[0]\right)^{-0.5}$ makes the quantization noise $\mathbf{q}'[n]$ to be uncorrelated with the unquantized observation vector $\mathbf{y}[n]$[3]. Let $\underline{\mathbf{y}} \triangleq [\mathbf{y}[N-1]^T\ \mathbf{y}[N-2]^T\ \ldots\ \mathbf{y}[0]^T]^T$. For the simplicity of the channel estimation part, $\mathbf{C}_{\underline{y}}$ is approximated as

$$\mathbf{C}_{\underline{y}} = \underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{p}}}\underline{\mathbf{H}}^H + N_o\mathbf{I} \approx \left(1/N_p(\rho_p^2|\mathcal{U}_P|K + \rho_i^2|\mathcal{U}_I|K) + N_o\right)\mathbf{I}, \tag{6.38}$$

where $\mathbf{R}_{\underline{\mathbf{p}}} = \mathbf{E}[\underline{\mathbf{p}}\underline{\mathbf{p}}^H]$, in which, $\underline{\mathbf{p}} = [\mathbf{p}[N-1]^T\ \mathbf{p}[N-2]^T\ \ldots\ \mathbf{p}[0]^T]^T$, where $\mathbf{p}[n]$ is an $M \times 1$ column vector whose k^{th} element is $p_k[n]$. Without the approximation, $MN \times MN$ matrix $\mathbf{C}_{\underline{y}}$ will be non-diagonal in general, which implies that $MN \times MN$ quantization noise covariance matrix will be non-diagonal. This will require taking the inverse of such a large matrix for LMMSE channel estimation as in [30] for frequency selective channel, which is computationally exhaustive. However, the approximation error goes to zero as L grows large which can be shown similarly as performed for $\mathbf{C}_{\mathbf{r}}[0]$ in Appendix B.4. Such an approximation is also adopted in [23]. Taking the DFT of the quantized observation vector $\mathbf{v}[n]$, it is found that

$$\tilde{\mathbf{v}}[u] = G'\tilde{\mathbf{y}}[u] + \tilde{\mathbf{q}}'[u], \tag{6.39}$$

where $G' = 2/\sqrt{\pi}\left((\rho_p^2|\mathcal{U}_P|K + \rho_i^2|\mathcal{U}_I|I)/N_p + N_o\right)^{-0.5}$. (6.39) along with (6.35) implies that

$$\tilde{y}_m[u] = \rho_p G'\sqrt{N_p K}h_{m,f(u)}[u]e^{j\phi_{f(u)}[u]} + G'z_m[u] + p_m[u], \tag{6.40}$$

where $u \in \mathcal{U}_P$ and $p_m[u]$ is the mth element of the DFT of $\mathbf{p}[n]$. It can be seen from (6.40) that the observation $\tilde{y}_m[vK+k-1]$, when the noise terms are omitted, is a phase rotated and scaled version of the channel coefficient $\tilde{h}_{m,k}[vK+k-1]$ for user k, sampled with a sampling period of K, as $f(vK+k-1) = (vK+k-1 \bmod K)+1 = k$. These samples will be denoted by $\check{h}_{m,k}[v] \triangleq \tilde{h}_{m,k}[vK+k-1]$. According to the Nyquist sampling theorem, if N_p satisfies

$$N_p \geq KL, \tag{6.41}$$

it is possible to obtain the channel coefficients without any aliasing. Again for the simplicity of the channel estimation part, the covariance matrix of $\mathbf{q}'[n]$ will be approximated as a diagonal matrix $(2 - 4/\pi)\mathbf{I}$, with approximation error going to

[3] Here, the quantizer inputs are again assumed to be Gaussian due to the same reasoning discussed in Footnote 1 of Section 6.3.

zero as L grows large for low oversampling rates and $\rho_p^2 \approx \rho_i^2$ as discussed in Appendix B.4. Under this approximation, the LMMSE estimate for the channel coefficient $\tilde{h}_{m,k}[vK+k-1]$, namely $\tilde{h}_{m,k}[vK+k-1]^*$, is found as

$$\tilde{h}_{m,k}[vK+k-1]^* = \frac{e^{-j\phi_k[vK+k-1]}\tilde{y}_m[vK+k-1]}{\rho_p G' \sqrt{N_p K}\left(1 + N_o/(\rho_p^2 K) + P_q/\left(\rho_p^2 K\left(G'\right)^2\right)\right)}, \quad (6.42)$$

where the quantization distortion variance $P_q = \mathbf{E}[|p_m[u]|^2] \approx 2 - 4/\pi$. Here, the channel estimation error σ_e^2 can also be found as

$$\sigma_e^2 = 1 - \frac{1}{\rho_p \sqrt{N_p K}\left(1 + N_o/(\rho_d^2 K) + P_q/\left(\rho_d^2 K\left(G'\right)^2\right)\right)}. \quad (6.43)$$

The parameter σ_e^2 will be used in the performance analysis of the investigated uplink system model in this work for imperfect CSI and the results from the analysis will be compared to the simulated results. As can be noted in (6.42), for the kth user, we only have the channel coefficients estimates for $\check{h}_{m,k}[v] \triangleq h_{m,k}[vK+k-1]^*$, sampled with a period of K. To obtain the remaining channel coefficients, $\check{h}_{m,k}[v]$ can be upsampled by K. There are many possibilities to upsample $\check{h}_{m,k}[v]$. The one adopted in this study is the spline interpolation [65].

The most important parameter through which the ACI is taken into account in the proposed channel estimation method is through the factor G' in (6.42). By definition, G' decreases with increasing total ACI power $\rho_i^2|\mathcal{U}_I|I/N_p$. The distortion caused by the quantization can be regarded to have two components, the additive quantization noise distortion $\tilde{\mathbf{q}}'[u]$ and the magnitude distortion G' in (6.39). Under the aforementioned approximations, the power of the additive quantization noise $\tilde{\mathbf{q}}'[u]$ does not change with the ACI power. However, as G' decreases with increasing ACI power, the power of the signal part $G'\tilde{\mathbf{y}}[u]$ in (6.39) diminishes, resulting in a reduced signal power compared to the quantization noise power. Therefore, a worse estimation error performance can be expected. This can also be interpreted from (6.43). The channel estimation error variance σ_e^2 in (6.43) increases as G' decreases with increasing ACI power. Regarding how the estimator combats with the degredation due to ACI can be inferred from (6.42). Neglecting the $P_q/(\rho_p^2 K(G')^2)$ term in the denominator in (6.42), it can be stated that as the ACI power is increased, which in turn decreases G' and reduces the signal component $G'\tilde{\mathbf{y}}[u]$ in (6.39), the estimator tries to cancel this effect by multiplying the observation by $1/G'$ (note the G' factor in the denomi-

nator in (6.42)). However, such a normalization (multiplication by $1/G'$ when G' is smaller than 1) results in the enhancement of the quantization noise $\tilde{\mathbf{q}}'[u]$, thus the cancellation of the magnitude distortion G' should be balanced with the quantization noise enhancement. This balancing is performed through the $P_q/\rho_p^2 K(G')^2$ factor in the estimator in (6.42).

6.5 ADCs with Higher than One-Bit Resolution

In this section, the details for the performance analysis for quantizers with more than one-bit resolution is presented. To begin with, we define the set of quantizer output values $\mathcal{L} = \{\ell_0, \ell_1, \ldots, \ell_{L'-1}\}$, where $L' = 2^q$ is the number of possible quantizer output values q being the number of ADC bits. Moreover, the quantization thresholds can also be characterized by the set $\mathcal{B} = \{b_0, b_1, \ldots, b_{L'}\}$, where $-\infty = b_0 < b_1 < \cdots < b_{L'} = \infty$. The quantization function $\mathcal{Q}(.)$ is a point in the function space $\mathbb{C}^M \to \Upsilon^M$, where $\mathbb{C}^M$ denotes the complex vector space of dimension M and $\Upsilon = \mathcal{L} \times \mathcal{L}$ is the set of possible quantizer output values (the cartesian product $\mathcal{L} \times \mathcal{L}$ represents the combination of the outputs of the pair of ADCs quantizing the real and imaginary parts of the received signals separately). The ith element of the quantizer output, namely $\mathcal{Q}(\mathbf{x})_i$, where $M \times 1$ quantizer input vector $\mathbf{x}$ can be expressed as

$$\mathcal{Q}(\mathbf{x})_i = \left(\ell_{f'(\mathrm{Re}(x_i))}, \ell_{f'(\mathrm{Im}(x_i))}\right), \tag{6.44}$$

where $f'(\mathrm{Re}(x)) = d \in \{0, 1, \ldots, L'-1\}$ which satisfies $b_d \leq \mathrm{Re}(x) < b_{d+1}$. Similarly $f(\mathrm{Im}(x)) = c \in \{0, 1, \ldots, L'-1\}$ which satisfies $b_c \leq \mathrm{Im}(x) < b_{c+1}$. As an example, the possible quantizer output values $\ell_i = \Delta(i - L'/2 + 1/2)$, $i = 0, 1, \ldots, L'-1$, whereas the quantization thresholds $b_i = \Delta\left(i - \frac{L'}{2}\right)$, $i = 1, 2, \ldots, L'-1$ for a uniform midrise quantizer ($b_0 = -\infty, b_{L'} = \infty$ as previously specified).

An important point in the design of the quantizer is the selection of the step size Δ. In fact, AGC will dynamically adjust the gain of the input signal to ADC according to the received signal power in order that it fits the input signal range of the ADC. This will correspond to the approach in this study in which the step size is selected according to the received signal power levels, which is assumed to stay nearly the same over a coherence interval. This will result in a fixed step size during a coherence interval,

enabling a tractable analysis.

There are two main considerations in the design of the step size Δ. If the step size is selected to be small for the average received signal power level, the probability that the input signal is clipped will be high and cause a distortion, referred to as *overload* distortion. On the other hand, if a large step size is preferred to avoid clipping or overload distortion, this will result in a granular distortion, causing a large range of input signal level to be mapped to the same level. Therefore, step size should be selected properly to balance the aforementioned granular and overload distortions. The amount of the two distortions will affect the validity of the assumptions in the performance analysis, as will be discussed in the subsequent parts of this section.

To begin with the analysis, matrix $\mathbf{A}$ in (6.12) should be evaluated. According to Bussgang decomposition, $\mathbf{A} = \mathbf{C_{d[n]r[n]}C^{-1}_{r[n]}}$, where $\mathbf{C_{d[n]r[n]}} = \mathbf{E[d[n]r[n]^H]}$ and $\mathbf{C_{r[n]}} = \mathbf{E[r[n]r[n]^H]}$. For the example case of midrise uniform quantizer with Gaussian inputs[4] [25],

$$\mathbf{A} = \frac{\Delta}{\sqrt{\pi}}\text{diag}\left(\mathbf{C}_r[0]\right)^{-0.5} \times \sum_{i=1}^{2^q-1} \exp\left(-\Delta^2\left(i-2^{q-1}\right)^2 \text{diag}\left(\mathbf{C}_r[0]\right)^{-0.5}\right). \quad (6.45)$$

As $\mathbf{C}_r[0]$ in (6.45) can be found using Proposition 2, matrix $\mathbf{A}$ can be calculated using (6.45) for multi-bit quantizer case. What remains is the calculation of the covariance matrix of the quantization noise $\mathbf{C_{q[u]}}$. The difficulty with the calculation of this matrix stems from the fact that there is no closed form expression for the relation between the quantizer input and output covariance matrices for multi-bit quantizers as for the one-bit quantizer in (6.19), which was referred to as the arcsine law. However, a diagonal approximation can be made, for which all non diagonal entries of the covariance matrix $\mathbf{C_d}[0]$ are assumed to be zero and the mth diagonal entry of $\mathbf{C_d}[0]$, namely $\mathbf{E}\left[|d_m[n]|^2\right]$, can be found as follows:

$$\begin{aligned}\mathbf{E}\left[|d_m[n]|^2\right] &= 2\sum_{i=0}^{L'-1} \ell_i^2 Pr\left(b_i \leq \text{Re}(r_m) < b_{i+1}\right) \\ &= 2\sum_{i=0}^{L'-1} \ell_i^2 \left(\Phi\left(\sqrt{2}b_{i+1}/\sigma_{r_m}\right) - \Phi\left(\sqrt{2}b_i/\sigma_{r_m}\right)\right), \end{aligned} \quad (6.46)$$

[4] The multi-bit quantizer input is also assumed to be Gaussian, which is accurate owing to the same reasoning discussed in Footnote 1 of Section 6.3.

where $\sigma_{r_m}^2$ is the mth diagonal element of $\mathrm{diag}(\mathbf{C_r}[0])_{m,m}$ corresponding to the quantizer input variance at the mth antenna and $Pr(.)$ denotes the probability of the event in its operand. In (6.46), it is assumed that the quantizer input has a Gaussian distribution[4]. We will denote the diagonal matrix whose diagonal entries are equal to the diagonal entries of $\mathbf{C_d}[0]$ as $\mathbf{C_d^{diag}}[0]$. After finding $\mathbf{C_d^{diag}}[0]$ from (6.46), the diagonal approximation for the $\mathbf{C_q}[0]$, namely $\mathbf{C_q^{diag}}[0]$, can be found from (6.18) as

$$\mathbf{C_q^{diag}}[0] = \mathbf{C_d^{diag}}[0] - \mathbf{A}\mathrm{diag}(\mathbf{C_r}[0])\mathbf{A}^H. \tag{6.47}$$

The covariance matrices of quantization noise for nonzero lags, namely $\mathbf{C_q^{diag}}[m]$, $m \neq 0$, are also assumed to be zero for multi-bit quantizers, which means that the correlation in time for the quantization noise is assumed to be zero. This assumption fails to be valid for very low ADC resolutions [26] or for high oversampling rates, as discussed in Appendix B.4, yet, it provides progressively more accurate results as the number of quantization bits is increased [26], when clipping or overload distortion occurs with low probability. To ensure this, we will choose the step size Δ small enough as will be discussed shortly. After finding $\mathbf{C_q^{diag}}[0]$, $\mathbf{C}_q[u]$ for the diagonal approximation case, which is referred to as $\mathbf{C}_q^{\mathrm{diag}}[u]$, can be found using Proposition 3 as follows:

$$\begin{aligned}\mathbf{C}_q^{\mathrm{diag}}[u] &= \mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[m]\right\} + \mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[m]\right\}^* - N\mathbf{C_{\tilde{q}}}[0] \\ &= N\mathbf{C_{\tilde{q}}}[0],\end{aligned} \tag{6.48}$$

as it is assumed for multi-bit quantizers that $\mathbf{\Gamma}[m] = (N-m)\mathbf{C_q}[m] \approx (N-m)\mathbf{C_q^{diag}}[m] = 0$ for $m \neq 0$. Then, $\mathbf{C_q^{diag}}[u]$ can be used to find the SINDR expression in Proposition 4, which can be employed to find the error-rate performance using (6.29). To ensure that the overload distortion is negligible, we adjust the step size Δ as follows:

$$\Delta = 2A_{max}/L', \tag{6.49}$$

where the maximum quantizer output level A_{max} is adjusted as $A_{max} = \sqrt{G/2}\left(1 - \Phi\left(P_c/2\right)\right)$, where G is as defined in Proposition 5, which corresponds to the average received power and P_c is the desired probability that a clipping occurs. Obviously, the received signal is also assumed to be Gaussian distributed in the adjustment of the step size, which is an accurate assumption according to Footnote 1. The clipping probability will be chosen as a small number, owing to its impact on the

validity of the diagonal approximations involved in the analysis. Imperfect CSI case for multi-bit quantizer is left for future work.

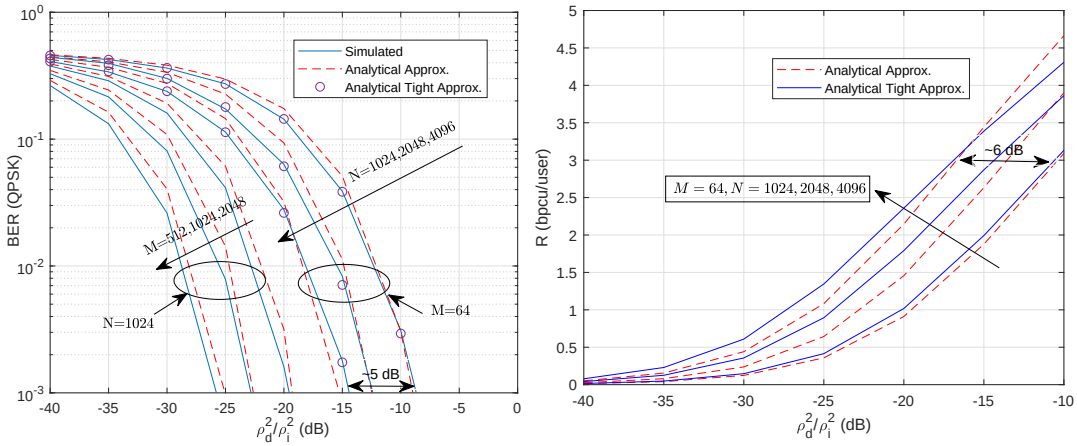
To see the effect of oversampling, number of ADC bits and number of antennas on the SINDR for the multi-bit quantizer case, the parameter G in (6.31) can be found using (6.45) as

$$G = \frac{\Delta}{\sqrt{\pi}} (\lambda)^{-0.5} \sum_{i=1}^{2^q-1} \exp\left(-\Delta^2 \left(i - 2^{q-1}\right)^2 (\lambda)^{-0.5}\right), \tag{6.50}$$

where $\lambda = (|\mathcal{U}_D| K \rho_d^2 + |\mathcal{U}_I| I \rho_i^2) / N + N_o$. The SINDR expression in (6.31) will be the same for the multi-bit quantizer case except that the parameter G in (6.31) is found according to (6.50) and the quantization noise variance Var$[q_k[u]]$ will also be another constant less than $2 - 4/\pi$, which is decreasing with the number of ADC resolution bits, but will not change with the number of antennas or the oversampling rate. As mentioned before, increasing N while $|\mathcal{U}_D|$ and $|\mathcal{U}_I|$ are fixed corresponds to an increase in the oversampling rates. In such case, G is increased as λ decreases with N. Therefore, the same discussion that SINDR $\gamma_k'[u]$ increases with G or the oversampling rates for one-bit ADC also applies for the multi-bit quantizer. Moreover, due to the $(M - K)$ factor in (6.31), it is also possible to increase SINDR by increasing the number of antennas M. Furthermore, since the step size Δ decreases when the number of ADC bits is increased, this corresponds to an increase in G according to (6.50), and a decrease in the quantization noise variance Var$[q_k[u]]$, which in turn results in an increase in the SINDR $\gamma_k'[u]$ in (6.31). Therefore, it can be stated that SINDR will increase when the oversampling rates, number of antennas and ADC bits are increased for the multi-bit quantizer case, in line with the intuition.

6.6 Simulation Results

For the simulations, unless otherwise stated, the number of receive antennas is $M = 64$, while $K = I = 4$. Some other parameters are $N = 1024$, $L = 10$ and $\mathcal{U}_D = \{N - 150, N - 149, \ldots, -1, 1, 2, \ldots, 150\}$ ($|\mathcal{U}_D| = 300$), while $\mathcal{U}_I = \{250, 251, \ldots, 549\}$ ($|\mathcal{U}_I| = 300$), which makes the oversampling rates $\beta_D = \beta_I \approx 3.41$. The data symbols are QPSK modulated. Subcarrier spacing is 15 kHz as in LTE, with a transmission bandwidth of $|\mathcal{U}_D|/(NT_s) = 4.5$ MHz for the desired channel, which does



(a) BER vs. SIR (ρ_d^2/ρ_i^2).

(b) R vs. SIR (ρ_d^2/ρ_i^2).

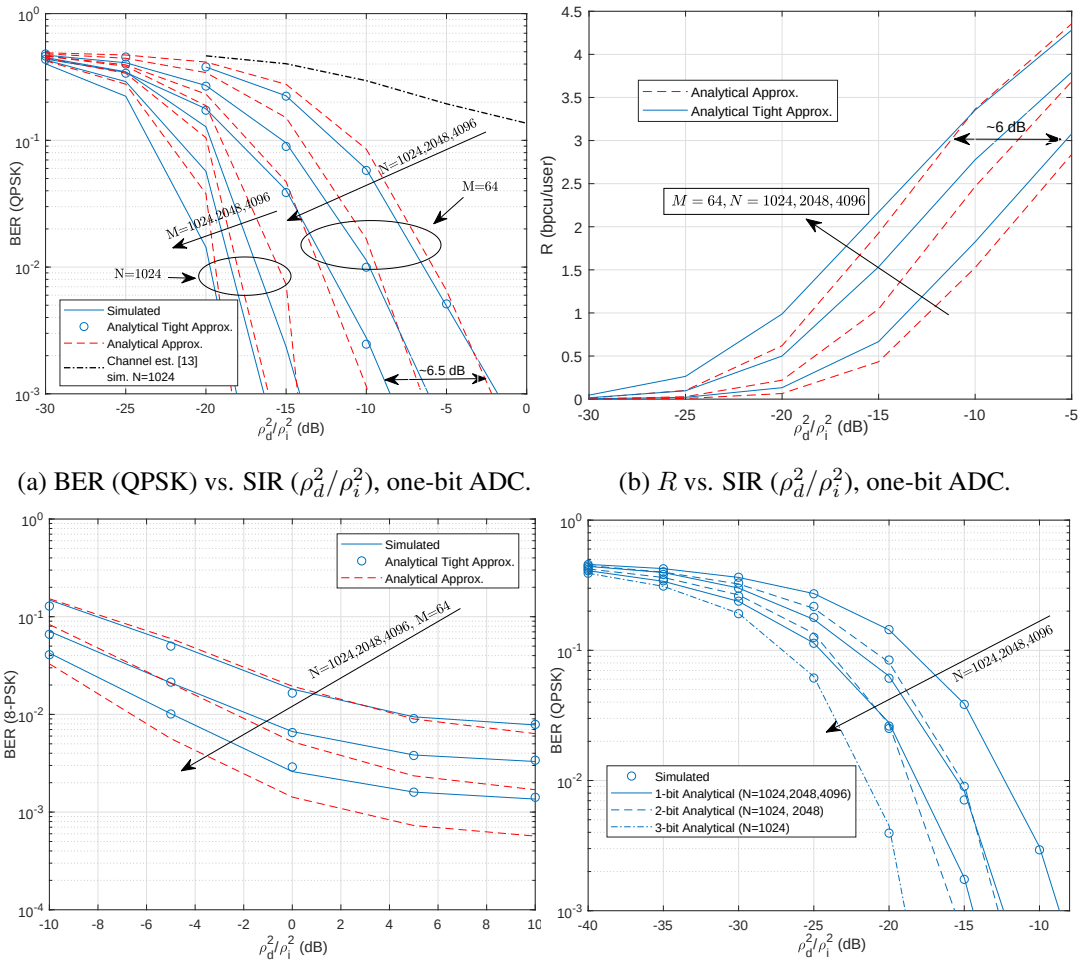
Figure 6.3: BER vs. SIR in (a) and R vs. SIR in (b), $M = 64$, $K = I = 4$, 1-bit ADC, perfect CSI.

not change with the sampling rate. Furthermore, the noise variance parameter N_o is normalized such that $\rho_d^2/N_o = 4$ dB to take ρ_d^2 as unity. The type of the multi-bit quantizers is uniform midrise, for which $P_c = 1\%$. The power delay profile is taken as uniform, that is, $p[l] = 1/L$ for $0 \leq \ell < L$. In the plots, the analytical curves for which the SINDR calculation is made based on Proposition 4 are referred to as "Analytical Tight Approx.", indicated with dashed lines. Moreover, the curves for which the SINDR is calculated based on Proposition 5 are named "Analytical Approx.", indicated with circles in Fig. 6.3a and with solid lines in Fig. 6.3b. As the first case, we change the block length N, when all other parameters are fixed (except L which should be directly proportional to N) for the perfect CSI condition. The BER vs signal-to-interference ratio (SIR or ρ_d^2/ρ_i^2) curves are presented in Fig. 6.3a. Note that a single ρ_d^2/ρ_i^2 for each data point does not imply that the average received power for every band (desired or interference band) or subcarrier/user is the same for a given channel realization due to (6.14). In Fig. 6.3a, the simulated values are indicated with solid lines. As can be noted in the three curves grouped as $M = 64$ curves on the right hand side of Fig. 6.3a, the analytical calculations based on Proposition 4 and (6.29) are in good agreement with the simulated values. Moreover, the approximate analytical curves based on Proposition 5 generally follow the simulated curves. In addition, we see that increasing the oversampling rate (equivalently increasing the block length while the number of occupied subcarriers is fixed) and the

number of antennas M are useful to combat ACI. We can observe up to 5 dB SIR gain by increasing the oversampling rate from $\beta_d \approx \beta_i = 3.41$ $(N = 1024, L = 10)$ to $\beta_d = \beta_i \approx 13.65$ $(N = 4096, L = 40)$ when the SIR levels to achieve a target BER of 10^{-3} is considered. Significant SIR gains are also observed in Fig. 6.3a when M is increased as expected from the analysis.

For the same simulation setting, the ergodic capacity in terms of bits per channel use (bpcu) per user calculated using (6.25) are plotted in Fig. 6.3b. An SIR gain more than 5 dB is observed with increasing oversampling rate when the SIR levels to achieve an ergodic capacity of 3 bpcu per user are compared. Moreover, it should be noted that the approximate analytical curve based on Proposition 5 is close to the tight approximation curve in Proposition 4 for $N = 1024$, a relatively low oversampling rate case, verifying that the approximation in Proposition 5 is accurate for low oversampling rates and when L is large[5]. Furthermore, it can be noticed that the approximate curve based on Proposition 4 yields higher BER for low SIR values or lower BER for high SIR values. The reason for this is as follows. For low SIR values, it can be shown that each element of $\mathbf{C_q}[m]$ will be the samples of an aliased sinc pulse (m being the sample index) centered around $m = 0$, all samples being real valued as $\mathcal{U}_D$ is symmetrical around the zeroth subcarrier. Since the values of the tails of the sinc pulse is much lower than that of its main lobe, it is reasonable to assume that $\mathbf{C_q}[m] \approx 0$ when $|m| > N/|\mathcal{U}_D| \triangleq W$, as $2N/|\mathcal{U}_D|$ is the null-to-null bandwidth of the sinc pulse). Therefore, $\mathbf{\Gamma}[m] \approx 0$ for $|m| > W$. As $\mathbf{C_q}[m] = \mathbf{C_q}[-m]$, $\mathbf{\Gamma}[m] = \mathbf{\Gamma}[-m]$, and $|\mathcal{U}_D| \gg 4$, $\mathrm{DFT}_{N,u}\{\mathbf{\Gamma}[m]\} \approx \sum_{m=-W}^{W} \mathbf{\Gamma}[m] e^{-j2\pi mu/N} = \sum_{m=-W}^{W} \mathbf{\Gamma}[m] cos(2\pi mu/N) > \mathbf{\Gamma}[0]$. Since the approximation assumes that $\mathbf{\Gamma}[m] = 0$ for $m \neq 0$ and $\sum_{m=-W}^{W} \mathbf{\Gamma}[m] > \mathbf{\Gamma}[0]$, the quantization noise covariance matrix calculated using Proposition 3 under this assumption has lower values compared to its exact version, resulting in a higher SINDR calculation than the exact values. For the low SIR case, it can be shown that each element of $\mathbf{C_{q[u]}}$ will mostly be concentrated inside the interfering band (for $u \in \mathcal{U}_I$) and the quantization noise in the desired band is due to the tails of sinc^2 pulses, making the variance of the quantization noise in the desired band limited less than the assumed

[5] In fact, R' should be compared to a slightly modified version of R in (6.25), for which the outer expectation is taken inside the logarithm, but the values obtained for this version are very similar to the values obtained for R in (6.25), thus not shown in Fig. 6.3b for the simplicity of the plot.



(a) BER (QPSK) vs. SIR (ρ_d^2/ρ_i^2), one-bit ADC.

(b) R vs. SIR (ρ_d^2/ρ_i^2), one-bit ADC.

(c) BER (8-PSK) vs. SIR (ρ_d^2/ρ_i^2), one-bit ADC.

(d) BER (QPSK) vs. SIR (ρ_d^2/ρ_i^2), multi-bit ADC.

Figure 6.4: Performance plots for imperfect CSI, 1-bit ADC in (a),(b),(c) and multi-bit ADC in (d).

quantization distortion value. The poor performance for the low SIR case is mostly due to the magnitude distortion caused by the matrix $\mathbf{A}$ in (6.12).

Simulations for the imperfect CSI case are also carried out. The pilot sequence length is taken to be N whereas the set of subcarriers for pilot signals is selected as $\mathcal{U}_P = \mathcal{U}_D \cup \{N-152, N-151, 151, 152\}$. The noise spectral density N_o in the channel estimation phase is also normalized such that $\rho_p^2/No = 4$ dB to take $\rho_p = 1$. The BER vs SIR plots are presented in Fig. 6.4a. As can be noted in Fig. 6.4a, the analytical BER curves obtained based on the SINDR calculation in Proposition 4 are very close to the simulated results. Moreover, while the approximate analytical curve is not as close to the simulated values as the "Analytical Tight Approx." curves, it can follow

the error rate curves in general. Moreover, it can also be deduced from Fig. 6.4a that the 5 dB SIR gain achieved with oversampling for the perfect CSI case can also be attained under imperfect channel knowledge when the error rates to achieve a BER value of 10^{-3} is considered. In fact, it is even more than 5 dB (about 6.5 dB) for imperfect CSI. This is because oversampling also enhances channel estimation quality which in turn results in a better error rate performance even further for the one-bit quantized system with imperfect CSI. Comparing the perfect and imperfect CSI BER curves in Fig. 6.3a and Fig. 6.4a, we observe about 6.5 dB SIR loss due to channel estimation error. This is not an unexpected value, as it is known that the normalized channel estimation mean squared error converges to -4.4 dB for infinite training power with one-bit ADCs [30], resulting in a similar SIR loss according to (6.31). The remaining 2.1 dB loss is due to the finite training power, which is close to the SNR loss of about 2 dB for a MIMO setting with infinite ADC resolution [66]. Moreover, to compare the proposed channel estimation algorithm with an existing method of comparable complexity in [23], which neither considers the effect of ACI nor employs oversampling for channel estimation, we made simulations to obtain the BER performance of our system with $N = 1024$ which uses the channel estimates obtained with the channel estimation method in [23]. The corresponding BER curve is labeled as “Channel Est. [13] Sim. N=1024" in Fig. 6.4a. As can be noted, a significant performance loss is observed if the channel estimation method in [23] is used instead of our method.

In addition to the BER curves, the ergodic capacity curves for imperfect CSI are also presented in Fig. 6.4b. As can be noted in Fig. 6.4b, more than 5 dB SIR advantage can be obtained with temporal oversampling when the SIR levels to achieve a target ergodic rate per user value of 3 bpcu/user are compared. Moreover, the “analytical approximate" curve can generally follow the “tight approximation" curve. The reason for the approximate ergodic capacity curve is not as close to the tight approximation curve for $N = 1024$ as in the perfect CSI case is due to the additional approximation error stemming from the assumptions involved in the proposed channel estimation scheme, distorting the orthogonality of the channel estimates and the estimation errors.

Regarding the performance with higher-order modulations, BER vs. SIR plots for

8-PSK and imperfect CSI are presented in Fig. 6.4c. As can be expected, worse BER performance is observed compared to QPSK. An error floor is observed since quantization and thermal noise exist even if the interference power is zero (infinite SIR). Moreover, a significant BER performance advantage is obtained with oversampling. The tight approximation based on Proposition 4 closely approximates the simulated values while the approximate curves based on Proposition 4 provides accurate values for low oversampling rates as expected from the discussion in Appendix B.4.

The error rate curves are also plotted for multi-bit quantizers (up to 3-bits) in Fig. 6.4d. As can be noted in Fig. 6.4d, the simulated values are very close to the analytical BER curves which are based on the SINDR calculation in Proposition 4. In Fig. 6.4d, the oversampling rate increases from 1024 to 4096 (L from 10 to 40) towards left for all quantization resolutions (1 bit to 3 bits). However, it should be noted that for 2-bit quantizer, there are two cases, either $N = 1024$ and $N = 2048$, while there is only the BER curve for $N = 1024$ for 3-bit quantizer. As can also be noted in Fig. 6.4d, the simulated values are in perfect agreement with the analytical values, even for the 2-bit quantizer, thus, it can be stated that the assumption of uncorrelated quantization noise in time is accurate unless the ADC resolution is as low as one bit (the quantization noise correlation in time is taken into account for the one-bit ADC case).

In order to make fair comparisons between the cases presented in Fig. 6.4d, we will try to equate the power consumptions of the various quantization resolution and oversampling rate cases. It is assumed that the power consumption of an ADC is proportional to 2^q, that is, the power consumption is doubled for single bit addition. This assumption is verified to be accurate in various studies [7, 8]. It is also assumed that ADC power consumption grows linearly with oversampling rate as in [8].

Equating the power consumptions, the first cases to be compared are 1-bit ADC with $N = 2048$ and 2-bit ADC with $N = 1024$. As can be noted in Fig. 6.4d, ACI supression in the 1-bit ADC with $N = 2048$ case is better compared to the 2-bit ADC with $N = 1024$ case for BER values higher than 10^{-2}, while the 2-bit ADC with $N = 1024$ case is slightly better for BER values lower than 10^{-3}. Therefore, it can be stated that 1-bit ADC with $N = 2048$ is preferable over a 2-bit ADC with $N = 1024$, as it provides better BER values and the complexity of a 1-bit ADC is significantly lower

(there is no need for an AGC unit in a 1-bit ADC, die area is doubled with every bit increase for flash ADCs; and component matching requirements are also doubled with every bit increase for flash, sucessive approximation (SAR) or pipelined ADCs [67]). Moreover, the time it takes to complete a conversion (conversion time) is also doubled with every bit of increase for integrating ADCs, while the conversion time scales linearly with number of bits for SAR or pipelined converters [67]. Therefore, by oversampling with a 1-bit ADC, while we achieve better error rate performance than 2-bit ADCs, when total power consumptions are kept equal, we also have advantages regarding ADC complexity and conversion time.

We can also compare other two cases, one is the performance of 1-bit ADC with $N = 4096$ and the other is 2-bit ADC with $N = 2048$, as their power consumptions are equal. We see from Fig. 6.4d that their performances are nearly equal for BER values lower than 10^{-2}, while the 2-bit ADC with $N = 2048$ case has about 1.5 dB SIR advantage for the BER value of 10^{-3}. The design engineer should be considering whether it is worth to have 1.5 dB SIR advantage to use 2-bit ADCs, which have the aforementioned disadvantages regarding implementation complexity and conversion time compared to the 1-bit ADCs.

The remaining performance comparison is between 3-bit ADCs with $N = 1024$ and 2-bit ADCs with $N = 2048$, as their power consumptions are equal. We can see that we have about 4 dB SIR advantage with 3-bit ADC. However, again, it should also be considered that such an SIR gain does not come for free, as 2-bit ADCs are much more advantageous compared to 3-bit ADCs in terms of die area, component matching circuitry and conversion time, thus 2-bit ADCs with oversampling can be a choice compared to 3-bit ADCs despite the SIR disadvantage.

6.7 Conclusions

In this work, we have presented a performance analysis for an uplink massive MIMO-OFDM system with low-resolution oversampling ADCs under frequency selective fading in an interfering adjacent channel interference scenario for perfect or imperfect receiver CSI. The analysis arrived at two important expressions, one of which gives

very accurate results but limited insights, the other giving noticeable approximation errors but much clearer insights into the dependence of system performance on the system parameters. We have shown both with analysis and simulations that adjacent band interference can be suppressed by increasing the number of antennas or the oversampling rate. Moreover, we discussed whether to use lower-resolution ADCs with higher oversampling rates or higher-resolution ADCs with lower oversampling rates comparing their error rate performances while their power consumptions are equated.

CHAPTER 7

A REDUCED COMPLEXITY UNGERBOECK RECEIVER FOR QUANTIZED WIDEBAND MASSIVE SC-MIMO

7.1 Motivation and Related Work

In this chapter, we question whether we can design a detector with superior performance compared to the existing detectors in the literature for quantized single-carrier massive MIMO (SC-MIMO) with comparable complexity. We especially investigate whether such a detector can be designed without resorting to any overesampling in time. To this end, we propose a novel iterative receiver for quantized CP-free uplink wideband SC-MIMO for uncorrelated or correlated Rayleigh and Rician fading channels. This detector utilizes an efficient message passing algorithm based on Bussgang decomposition, reduced state sequence estimation and Ungerboeck factorization. In this way, it achieves remarkable complexity reduction and exhibits significant performance advantages compared to the existing quantized SC-MIMO receivers from the literature. We also derive linear minimum mean-square-error channel estimator for cyclic-prefix (CP) free SC-MIMO under frequency-selective channel.

For the proposed channel estimation (CE) algorithm in this chapter, we concentrate on a linear and low complexity method for quantized frequency-selective MIMO, which can work with single-carrier (SC) modulation. The reason to select SC over OFDM for the proposed CE and data detection algorithms is that SC is superior to OFDM for systems having nonlinearities, such as quantized MIMO [68, 69], owing to its lower peak-to-average power ratio (PAPR) [70] and robustness to carrier-frequency-offset (CFO) errors [71]. Having lower PAPR is critical for systems with non-linear elements [72, 73], where [73] demonstrates the advantages of SC over

OFDM for MIMO systems with nonlinear elements. Even for unquantized linear systems, there are many recent studies that motivate the use of SC especially for MIMO structures [72–74], owing to its advantages related to channel equalization and spectral efficiency. In addition to the aforementioned advantages, the SC framework in the study in this chapter does not require any CP in constrast to conventional multi-carrier modulation schemes such as OFDM.

Numerous channel estimation (CE) algorithms in quantized flat-fading MIMO are mentioned in the survey paper [75]. However, frequency-flat channel assumption is not practical for wideband transmission [23]. As quantization is a non-linear operation, the extension of flat fading CE techniques to frequency-selective channels is not straightforward. Therefore, there are many works proposing various CE algorithms for frequency-selective quantized MIMO [23, 30, 69, 76–81]. Among them, [76] proposes a CE method to estimate sparse frequency-selective channels for orthogonal frequency division multiplexing (OFDM) modulation. More recently, [23] proposed a CE technique for frequency-selective MIMO-OFDM, without any sparsity assumption on the channel. In [23], quantization noise is assumed to be independent and identically distributed. This assumption is only accurate when the number of channel taps or users are large. In contrast, [30] takes into account the correlation in quantization noise by deriving the linear minimum mean-square-error (LMMSE) channel estimate. Another study [77] proposes a low-complexity CE algorithm based on approximate message-passing, showing some performance improvement compared to LMMSE channel estimation. However, the aforementioned CE techniques [23,30,77] require OFDM and a cyclic-prefix (CP), which may decrease the spectral efficiency significantly if the number of channel taps L is large. The same OFDM or CP limitation also exists for the CE techniques in [69, 78–81]. In short, for all of the aforementioned CE methods, at least one of the following limitations exists: the requirement of OFDM or a CP [23,30,69,76–81], the requirement of a sparse channel [69,76,78–80].

Regarding data detection in quantized massive MIMO, there are also a vast number of studies in the literature, some of which are mentioned in the survey paper [75]. However, as mentioned before, frequency-flat channel assumption is not a practical assumption. Therefore, [49, 81–90] advocate various data detectors for quantized massive MIMO systems under frequency-selective fading. Among those work, [49,

81, 82] propose data detectors for quantized massive MIMO but they are limited to OFDM modulation, which requires a CP. They are also highly complex compared to the proposed detector in this chapter (see Table 7.1 for details). Although lower complexity detectors compared to the highly complex maximum *a posteriori* (MAP) detector in [49] are also proposed in the same study, they are shown to provide an inferior performance compared to a much lower complexity per subcarrier LMMSE data equalization method [88].

Owing to the aforementioned advantages of SC systems over OFDM, there are many studies proposing SC frequency-domain equalization (SC-FDE) detectors in quantized MIMO. To start with, [83] proposes a generalized approximate message-passing (GAMP) based detector. However, the number of nonlinear operations per iteration of the proposed receiver in [83] is $5MNO + PKN$, where O is a number between 80 and 100, P is the modulation order, M is the number of antennas, K is the number of users and N is the data packet length, which can be compared to the number of subcarriers N_c of the multi-carrier modulation schemes. As the number of antennas M in massive MIMO is large, $5MNO$ becomes a very large number. Moreover, the number of iterations for GAMP based methods to converge is typically about 10 iterations [82], whereas the proposed detector in this chapter will be observed to converge in about $I = 2$ iterations in most cases. This means a prohibitive complexity for the detector in [83] for massive MIMO. Moreover, [84] also advocates a GAMP based receiver, but its complexity grows with N^2, which can also be very high. Another SC-FDE and GAMP based detector is proposed in [85]. Nevertheless, the detector in [85] is limited to spatial modulation, which is not a commonly used technique. Lately, [86] proposed various iterative detectors with feasible complexity for quantized massive MIMO.

Despite being superior to OFDM for quantized MIMO, there is still a CP overhead in SC-FDE. Therefore, [87, 88, 91] have recently proposed detectors that can work without a CP for quantized single-carrier MIMO (SC-MIMO) under frequency-selective fading. However, their complexity is very high compared to the proposed detector in this chapter (see Table 7.1 for details).

More recently, a maximum-likelihood sequence estimator for CP-free one-bit wide-

Table 7.1: Comparison of existing works with the proposed items in this chapter.

	Channel Estimation		Data Detection										
	[13],[50],[55],[62],[64],[66],[71],[74],[107]	Proposed	[74]	[91]	[27]	[104]	[21]	[105]	[23]	[38]	[39],[65]	[67]	Proposed
Requires OFDM/CP	All	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
Requires channel sparsity	[55],[62],[64] [71],[107]	No	No	No	No	No	No	No	No	No	No	No	No
Complexity growth vs. (proposed)	N/A	N/A	N^2K^2 (NK^2)	N/A	N^3 (N)	$MNOI_m+$ $PKNI_m$ (NK^2PLI)	N^2MK $(NK^2$ $+NMK)$	$INKM+$ $INLKM$ (NK^2PLI)	NK^3 (NK^2)	MN^3KP (NK^2PL)	NP^L (NPL)	NK^3 (NK^2)	N/A
Further limitations	No	No	No	Inferior performance	No	Slow convergence (High I_m, O)	No	Requires spatial modulation	No	No	No	No	No

band massive SC-MIMO has been proposed in [90]. The computational complexity of the detector in [90] grows with P^L, resulting in excessive complexity for large L. In the same study, the necessity of a decision feedback equalization based reduced-state detector is also mentioned as a future work. Such a detector is proposed in this chapter. Through decision feedback equalization, the number of states in the data detection algorithm can be reduced from P^L down to P by making decisions about the past and future symbols according to the available observations. In this way, we reduce the detection complexity such that it grows linearly with L, instead of P^L.

All aforementioned schemes are compared and contrasted with the proposed detector in this chapter in Table 7.1 with advantageous (disadvantageous) properties shaded in green (red). As the computational complexity analysis for the related works is limited to only some of the system parameters, we are only able to compare the complexity growth of the proposed detector with respect to those parameters. For simplicity, we are taking the terms in the complexity growth expressions with the largest powers of the parameters in comparison as they will be determining the complexity growth when parameters used in the comparisons are large. The complexity growth associated with the proposed detector in this chapter are indicated inside parentheses in Table 7.1. Note that M and N can be as large as 256 [92] and 1024 [26] in massive MIMO, respectively. For detailed comparisons, the reader can refer to the aforementioned discussions. Among the detectors of comparable complexity, the ones proposed in [86] and [89] are the only detectors without limitations such as slow convergence or spatial modulation requirement. We prefer [86] over [89] as the benchmark algorithm, the reasons of which will be discussed in the sequel.

7.1.1 Contributions

The main contribution items of the study in this chapter are as follows:

- It is the first study to derive LMMSE channel estimator for CP-free quantized SC-MIMO.
- The proposed detector is one of the few detectors in the literature that can work without CP in quantized (one or multi-bit) massive SC-MIMO, which can be important for a feasible spectral efficiency. In the literature, the detectors working without CP has a complexity growth with P^L, whereas the complexity of the proposed detector grows linearly with L.
- The proposed detector is the first reduced state Ungerboeck-type detector with bidirectional decision feedback structure working in wideband MIMO even for the unquantized case.
- The proposed detector provides significant error-rate performance advantages over the benchmark detector [86] from the literature.

For benchmark detector selection, the ones with comparable complexity to our detector are [86] and [89]. We prefer [86] over [89] as the benchmark detector even if there is a CP overhead in [86] (thus it is spectrally inefficient). The reason is that the performance of the algorithm proposed in [89] may be inferior to the algorithm in the much recent study [86] due to the inter-carrier interference caused by the lack of CP in [89].

7.2 System Model

In this chapter, a frequency-selective single-cell uplink massive MIMO system with K single-antenna users and M receive antennas with low-resolution ADCs is examined. The unquantized received signal at the mth antenna can be written by replacing T by the sampling rate T_s in (2.2) in Chapter 2 as

$$r_m(t) = \sum_{\ell=0}^{L-1} \sum_{k=1}^{K} \sum_{n=1}^{N} h_{m,k}[\ell] x_{n,k} p_c(t - (n-1)T_s - \ell T_s) + w_m(t), \tag{7.1}$$

where $p_c(t) = sinc(t/T_s)$. It is assumed that the sampling rate of the receiver is the same as that of the transmitter. As $sinc(t/T_s) = 0$ for $t = nT_s$ for $n \neq 0$, the discrete-time received signal at the mth antenna, namely $y_m[n] = r_m(nT_s)$, can be expressed as follows:

$$y_m[n] = \sum_{k=1}^{K} \sum_{\ell=0}^{L-1} \sqrt{\rho_k[\ell]} h'_{m,k}[\ell] x_k[n-\ell] + w_m[n], \tag{7.2}$$

where $x_k[n] = x_{n+1,k}$, $n = 0, \ldots, N-1$, $w_m[n] = w_m(nT_s)$, L is the number of channel taps, and $h'_{m,k}[\ell] = h_{m,k}[\ell]/\sqrt{\rho_k[\ell]}$. The channel taps $h'_{m,k}[\ell]$ are generally assumed to be zero-mean unit variance circularly symmetric complex Gaussian (CSCG) random variables, corresponding to a Rayleigh fading scenario, and uncorrelated, that is, $\mathbf{E}[h'_{m_1,k_1}[\ell_1] h'_{m_2,k_2}[\ell_2]^*] = \delta[\ell_1 - \ell_2]\delta[k_1 - k_2]\delta[m_1 - m_2]$. Moreover, $\rho_k[\ell]$ is the power-delay profile (PDP) of the channel between the receive antennas and the kth user, satisfying $\sum_{\ell=1}^{L} \rho_k[\ell] = 1, \forall k$. The justification behind the uncorrelated channel assumption is discussed in Chapter 2. However, we will also examine the case where there is a spatial correlation between the channels observed by different antennas and Rician fading, in which the channels are consisting of a combination of a line-of-sight (LoS) path and a small-scale fading component, which is recently examined for massive MIMO in [93]. Let $\mathbf{h}_k[\ell]$ be the channel vector associated with the lth channel tap of the kth user, whose m^{th} element is $h'_{m,k}[\ell]$. For Rician fading and a spatially correlated case, $\mathbf{h}_k[\ell]$ can be modelled as a realization of the CSCG distribution with mean $\xi_k[\ell]$ and covariance matrix $\mathbf{R}_k[\ell]$ [93], which can be found as [24, 94]

$$\mathbf{R}_k[\ell] = \int_{-\pi}^{\pi} \varrho_k^{\ell}(\theta) \mathbf{q}(\theta) \mathbf{q}(\theta)^H d\theta \tag{7.3}$$

where $\varrho_k^{\ell}(\theta)$ is the angular power profile of the ℓth channel tap of user k and $\mathbf{q}(\theta)$ is the steering vector of the antenna array. For uniform-linear array (ULA) with half-wavelength antenna separation, $\mathbf{q}(\theta) = [1\ e^{j\theta} \ldots e^{j(M-1)\theta}]$, where $\theta = \pi sin(\phi)$, ϕ being the angle of arrival. For a uniform power distribution, restricted between $\theta_{k,1}^{\ell}$ and $\theta_{k,2}^{\ell}$, $\mathbf{R}_k[\ell]$ in (7.4) can be approximated for ULA with half-wavelength antenna separation as [94]

$$\mathbf{R}_k[\ell] \approx \mathbf{q}(\mu_{\theta,k}^{\ell}) \mathbf{q}(\mu_{\theta,k}^{\ell})^H \odot \mathbf{D}(\sigma_{\theta,k}^{\ell}), \tag{7.4}$$

where $\odot$ represents Hadamard product, $\mu_{\theta,k}^{\ell} = (\theta_{k,1}^{\ell} + \theta_{k,2}^{\ell})/2$, $\sigma_{\theta,k}^{\ell} = |\theta_{k,1}^{\ell} - \theta_{k,2}^{\ell}|$, $[\mathbf{D}(\theta)]_{(m,n)} = \operatorname{sinc}\left((m-n)\,\theta/(2\pi)\right)$. Here, $\theta_{k,1}^{\ell}$ and $\theta_{k,2}^{\ell}$ can be found from mean

arriving angle ϕ_k^ℓ and angular spread ς_k^ℓ as $\theta_{k,1}^\ell = \pi\sin(\phi_k^\ell - \varsigma_k^\ell/2)$ and $\theta_{k,2}^\ell = \pi\sin(\phi_k^\ell + \varsigma_k^\ell/2)$. Moreover, the mean vector can be found as $\xi_k[\ell] = \sqrt{\kappa_k[\ell]}\mathbf{q}(\mu_{\theta,k}^\ell)$ [93], where $\kappa_k[\ell]$ is the Rician factor, determining the relative power of the LoS path compared to the non LoS paths for the ℓth channel tap of user k. If a Rician fading scenario is considered, $\rho_k[\ell]$ is scaled with $1/(1+\kappa_k[\ell])$ to ensure unit received power from each user. Regarding the correlation of the channels of different users and channel taps, as the users and the scatterers resulting from different clusters (corresponding to different channel taps) are physically separated by multiple wavelengths in general, the channels of different users and channel taps can be well modelled as statistically uncorrelated [24, 93]. Thermal noise samples $w_m[n]$ are assumed as independent identically distributed (i.i.d) zero-mean CSCG random variables with variance N_o. Moreover, $x_k[n]$ in (7.2) is the transmitted symbol by user k at n^{th} time index, with average symbol energy $E_s = \mathbf{E}[|x_k[n]|^2], \forall k, n$. (7.2) can be rewritten as

$$\mathbf{y}[n] = \sum_{\ell=0}^{L-1} \mathbf{H}[\ell]\mathbf{J}[\ell]\mathbf{x}[n-\ell] + \mathbf{w}[n], \tag{7.5}$$

where $\mathbf{J}[\ell]$ is a $K \times K$ diagonal matrix, whose k^{th} diagonal is $\sqrt{\rho_k[\ell]}$, and $\mathbf{H}[\ell]$ is the MIMO channel matrix, whose element at its mth row and the kth column is equal to $h'_{m,k}[\ell]$. Moreover, $\mathbf{y}[n]$, $\mathbf{w}[n]$ and $\mathbf{x}[n]$ are vectors, whose m^{th} and k^{th} elements are equal to $y_m[n]$, $w_m[n]$, and $x_k[n]$, respectively. The quantized received signal can also be expressed as

$$\mathbf{r}[n] = \mathrm{Q}(\mathbf{y}[n]), \tag{7.6}$$

where $\mathrm{Q}(.)$ is the function mapping the input of the quantizer to its output. For 1-bit quantizer, $\mathrm{Q}(.) = \mathrm{sign}\left(\mathrm{Re}\left(.\right)\right) + j\mathrm{sign}\left(\mathrm{Im}\left(.\right)\right)$, $\mathrm{sign}(.)$ being the sign function.

7.3 LMMSE Channel Estimation for CP-free Quantized SC-MIMO

In this section, the expression for the LMMSE channel estimate for quantized CP-free SC-MIMO systems will be derived. In the channel estimation phase, we assume that each user transmits pilot signals simultaneously. Under such a scenario, the received signal in (7.5) can be reexpressed as

$$\underline{\mathbf{y}}^{(p)} = \left(\mathbf{X} \otimes \mathbf{I}_M\right)\underline{\mathbf{h}} + \underline{\mathbf{w}}, \tag{7.7}$$

where

$$\begin{aligned}
\underline{\mathbf{y}}^{(p)} &\triangleq \left[\mathbf{y}[0]^T\, \mathbf{y}[1]^T \cdots \mathbf{y}[\tau-1]^T\right]^T, \\
\underline{\mathbf{w}} &\triangleq \left[\mathbf{w}[0]^T\, \mathbf{w}[1]^T \cdots \mathbf{w}[\tau-1]^T\right]^T, \\
\mathbf{X} &\triangleq \left[\mathbf{X}_1\, \mathbf{X}_2 \cdots \mathbf{X}_K\right], \\
\underline{\mathbf{h}} &\triangleq \left[\left(\mathbf{h}^{(1)}\right)^T\, \left(\mathbf{h}^{(2)}\right)^T \cdots \left(\mathbf{h}^{(K)}\right)^T\right]^T, \\
\mathbf{h}^{(k)} &\triangleq \left[\left(\mathbf{h}^{(k)}[0]\right)^T\, \left(\mathbf{h}^{(k)}[1]\right)^T \cdots \left(\mathbf{h}^{(k)}[L-1]\right)^T\right]^T,
\end{aligned}$$

in which τ is training length, $\mathbf{h}^{(k)}[\ell]$ is the kth column of $\mathbf{H}[\ell]$, $[\mathbf{X}_k]_{(m+1,n+1)} \triangleq \sqrt{\rho_k[n]}x_k[m-n]$, where $x_k[m]$, $m = 0, 1, \ldots, \tau-1$, is the transmitted pilot of user k, and $n = 0, 1, \ldots, L-1$.

To obtain a linear and simple channel estimator, we utilize the Bussgang decomposition [32], through which a statistically equivalent linear operator can be found for any nonlinear function [31]. According to the Bussgang decomposition, $\underline{\mathbf{r}}^{(p)} = \mathrm{Q}(\underline{\mathbf{y}}^{(p)})$ can be written as

$$\underline{\mathbf{r}}^{(p)} = \underline{\mathbf{A}}^{(p)}\underline{\mathbf{y}}^{(p)} + \underline{\mathbf{q}}^{(p)}. \tag{7.8}$$

Note that (7.8) is also valid for oversampled signals [19, 26, 86]. The only difference that arise with oversampling is that the effect of the employed pulse-shape should be included in (7.5) while constructing the signal model as performed in [16]. However, this case is left as a future work. We denote the cross-covariance matrix between $\underline{\mathbf{y}}^{(p)}$ and $\underline{\mathbf{r}}^{(p)}$ by $\mathbf{C}_{\underline{\mathbf{y}}^{(p)}\underline{\mathbf{r}}^{(p)}}$ and the autocovariance matrix of $\underline{\mathbf{y}}^{(p)}$ by $\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}$. When $\underline{\mathbf{A}}^{(p)}$ is selected as $\underline{\mathbf{A}}^{(p)} = \mathbf{C}^H_{\underline{\mathbf{y}}^{(p)}\underline{\mathbf{r}}^{(p)}}\mathbf{C}^{-1}_{\underline{\mathbf{y}}^{(p)}}$, the distortion term $\underline{\mathbf{q}}^{(p)}$ in (7.8) is minimized. Equivalently, $\underline{\mathbf{q}}^{(p)}$ is made uncorrelated with $\underline{\mathbf{y}}^{(p)}$. In order to find the LMMSE channel estimator, two critical quantities that should be found are matrix $\underline{\mathbf{A}}^{(p)}$ and the autocovariance matrix of $\underline{\mathbf{q}}^{(p)}$, namely $\mathbf{C}_{\underline{\mathbf{q}}^{(p)}}$. We will find these terms for two different cases, one being the one-bit and the other being the multi-bit quantizer case. We denote $\underline{\mathbf{A}}^{(p)}$, $\mathbf{C}_{\underline{\mathbf{q}}^{(p)}}$ and the autocovariance matrix of $\underline{\mathbf{r}}^{(p)}$ for one-bit quantizer case by $\underline{\mathbf{A}}^{(p,1)}$, $\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}}$ and $\mathbf{C}^1_{\underline{\mathbf{r}}^{(p)}}$ and for multi-bit quantizer case by $\underline{\mathbf{A}}^{(p,m)}$, $\mathbf{C}^m_{\underline{\mathbf{q}}^{(p)}}$ and $\mathbf{C}^m_{\underline{\mathbf{r}}^{(p)}}$, respectively. In this case,

$$\begin{aligned}
\underline{\mathbf{A}}^{(p)} = \underline{\mathbf{A}}^{(p,1)}\chi_q + \underline{\mathbf{A}}^{(p,m)}(1-\chi_q),\ \ \mathbf{C}_{\underline{\mathbf{r}}^{(p)}} = \mathbf{C}^1_{\underline{\mathbf{r}}^{(p)}}\chi_q + \mathbf{C}^m_{\underline{\mathbf{r}}^{(p)}}(1-\chi_q), \\
\mathbf{C}_{\underline{\mathbf{q}}^{(p)}} = \mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}}\chi_q + \mathbf{C}^m_{\underline{\mathbf{q}}^{(p)}}(1-\chi_q),
\end{aligned} \tag{7.9}$$

where q is the number of quantizer bits and χ_q is an indicator function, defined as $\chi_q = 1$ if $q = 1$ and $\chi_q = 0$, otherwise. The matrices $\underline{\mathbf{A}}^{(p,1)}$, $\underline{\mathbf{A}}^{(p,m)}$, $\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}}$ and $\mathbf{C}^m_{\underline{\mathbf{q}}^{(p)}}$ are all found for one-bit or multi-bit quantizer cases in (C.1), (C.2), (C.7) and (C.8) in Appendix C.1. Now that they are found, the quantized signal $\underline{\mathbf{r}}^{(p)}$ can be found using (7.7), (7.8), (7.9) and (C.1) for one-bit or (7.7), (7.8), (7.9) and (C.2) for multi-bit quantizer case as

$$\underline{\mathbf{r}}^{(p)} = (\mathbf{B}\mathbf{X} \otimes \mathbf{I}_M)\,\underline{\mathbf{h}} + (\mathbf{B} \otimes \mathbf{I}_M)\,\underline{\mathbf{w}} + \underline{\mathbf{q}}^{(p)}, \tag{7.10}$$

where $\mathbf{B} = \chi_q \mathbf{B}_1 + (1-\chi_q)\mathbf{B}_m$, in which $\mathbf{B}_1$ and $\mathbf{B}_m$ are as found in Appendix C.1. Define the total effective noise at the quantizer output as $\underline{\mathbf{\Gamma}} \triangleq (\mathbf{B} \otimes \mathbf{I}_M)\,\underline{\mathbf{w}} + \underline{\mathbf{q}}^{(p)}$. Its covariance matrix $\mathbf{C}_{\underline{\mathbf{\Gamma}}}$ can be found using (C.7) and (7.10) for one-bit quantizer or using (C.8) and (7.10) for multi-bit quantizer as

$$\mathbf{C}_{\underline{\mathbf{\Gamma}}} = \mathbf{F} \otimes \mathbf{I}_M, \tag{7.11}$$

where $\mathbf{F} = N_o \mathbf{B}\mathbf{B}^H + \chi_q \mathbf{E}_1 + (1-\chi_q)\mathbf{E}_m$, $\mathbf{E}_1$ and $\mathbf{E}_m$ are as found in Appendix C.1. As effective noise covariance matrix is found, we can apply a whitening filter, $\mathbf{C}_{\underline{\mathbf{\Gamma}}}^{-1/2} = \mathbf{F}^{-1/2} \otimes \mathbf{I}_M$, to obtain

$$\underline{\mathbf{z}}^{(p)} \triangleq \mathbf{C}_{\underline{\mathbf{\Gamma}}}^{-1/2} \underline{\mathbf{r}}^{(p)} = (\mathbf{P}\mathbf{X} \otimes \mathbf{I}_M)\,\underline{\mathbf{h}} + \underline{\mathbf{n}}, \tag{7.12}$$

where $\mathbf{P} = \mathbf{F}^{-1/2}\mathbf{B}$ and $\underline{\mathbf{n}} = (\mathbf{P} \otimes \mathbf{I}_M)\,\underline{\mathbf{w}} + \mathbf{C}_{\underline{\mathbf{\Gamma}}}^{-1/2} \underline{\mathbf{q}}^{(p)}$, whose covariance matrix $\mathbf{C}_{\underline{\mathbf{n}}} = \mathbf{I}_{M\tau}$. To derive the LMMSE estimator, we also need to find whether $\underline{\mathbf{h}}$ and $\underline{\mathbf{n}}$ are uncorrelated. It has been shown in [30, Appendix A] that for any quantized LMMSE channel estimation based on Bussgang decomposition, the quantization distortion term, which is referred to as $\underline{\mathbf{q}}^{(p)}$ in this chapter, is uncorrelated with the channel, implying that $\underline{\mathbf{n}}$ is also uncorrelated with $\underline{\mathbf{h}}$ as $\underline{\mathbf{w}}$ is also uncorrelated with $\underline{\mathbf{h}}$. In this case, $\mathbf{C}_{\underline{\mathbf{z}}^{(p)}} \triangleq \mathbf{E}[\underline{\mathbf{z}}^{(p)} \underline{\mathbf{z}}^{(p)H}]$ and $\mathbf{C}_{\underline{\mathbf{z}}^{(p)}\underline{\mathbf{h}}} \triangleq \mathbf{E}\left[\underline{\mathbf{z}}^{(p)} \underline{\mathbf{h}}^H\right]$ can be found as

$$\mathbf{C}_{\underline{\mathbf{z}}^{(p)}} = (\mathbf{X}' \otimes \mathbf{I}_M)\,\mathbf{C}_{\underline{\mathbf{h}}} \big(\mathbf{X}' \otimes \mathbf{I}_M\big)^H + \mathbf{I}_{M\tau}, \quad \mathbf{C}_{\underline{\mathbf{z}}^{(p)}\underline{\mathbf{h}}} = (\mathbf{X}' \otimes \mathbf{I}_M)\,\mathbf{C}_{\underline{\mathbf{h}}}, \tag{7.13}$$

$\mathbf{X}' \triangleq \mathbf{P}\mathbf{X}$, $\mathbf{C}_{\underline{\mathbf{h}}}$ is the auto-covariance matrix of $\underline{\mathbf{h}}$. Consequently, the LMMSE channel estimate for CP-free quantized wideband SC-MIMO, namely $\hat{\underline{\mathbf{h}}}^{\text{LMMSE}}$, can be found using (7.13) as

$$\begin{aligned} \hat{\underline{\mathbf{h}}}^{\text{LMMSE}} &= \mathbf{C}^H_{\underline{\mathbf{z}}^{(p)}\underline{\mathbf{h}}} \mathbf{C}^{-1}_{\underline{\mathbf{z}}^{(p)}} \underline{\mathbf{z}}^{(p)} \\ &= \mathbf{C}_{\underline{\mathbf{h}}} (\mathbf{X}' \otimes \mathbf{I}_M)^H \left((\mathbf{X}' \otimes \mathbf{I}_M)\,\mathbf{C}_{\underline{\mathbf{h}}} \big(\mathbf{X}' \otimes \mathbf{I}_M\big)^H + \mathbf{I}_{M\tau} \right)^{-1} \underline{\mathbf{z}}^{(p)}. \end{aligned} \tag{7.14}$$

Note that in (7.14), inverting an $M\tau \times M\tau$ matrix may not be computationally feasible. In what follows, we obtain a much lower complexity channel estimator by taking the channel covariance matrix as identity. Such a selection of covariance matrix is appropriate when the channel is uncorrelated or when we do not have any prior knowledge on the channel covariance matrix thus take it as identity. The resulting estimate corresponds to the least-squares channel estimate for uncorrelated channel and high signal-to-quantization noise ratio case. To obtain this lower complexity estimator, corresponding to the exact LMMSE estimator for uncorrelated channels, we define $\mathbf{X}'' \triangleq \mathbf{X}' \otimes \mathbf{I}_M$. Then, (7.14) can be rewritten by replacing $\mathbf{C}_{\underline{\mathbf{h}}}$ with identity matrix as

$$\hat{\underline{\mathbf{h}}}^{\text{LMMSE}} = \mathbf{X}''^H \left(\mathbf{X}''\mathbf{X}''^H + \mathbf{I}_{M\tau}\right)^{-1} \underline{\mathbf{z}}^{(p)}. \tag{7.15}$$

Employing Woodbury matrix identity [95], (7.15) can be reexpressed as

$$\begin{aligned}
\hat{\underline{\mathbf{h}}}^{\text{LMMSE}} &= \left((\mathbf{X}'')^H \mathbf{X}'' + \mathbf{I}_{MKL}\right)^{-1} (\mathbf{X}'')^H \underline{\mathbf{z}}^{(p)} \\
&= \left(\left((\mathbf{X}')^H \mathbf{X}' \otimes \mathbf{I}_M\right) + \mathbf{I}_{MKL}\right)^{-1} (\mathbf{X}'')^H \underline{\mathbf{z}}^{(p)} \\
&= \left(\left(\mathbf{X}^H\mathbf{P}^H\mathbf{P}\mathbf{X} + \mathbf{I}_{KL}\right)^{-1} \otimes \mathbf{I}_M\right) (\mathbf{P}\mathbf{X} \otimes \mathbf{I}_M)^H \underline{\mathbf{z}}^{(p)}.
\end{aligned} \tag{7.16}$$

Taking the inverse of a $KL \times KL$ matrix in (7.16) is much less complex than taking the inverse of an $M\tau \times M\tau$ matrix in (7.14), as $\tau \geq KL$ in general (for orthogonal pilot assignment to users $\tau \geq KL$ [23]). This complexity, along with the complexity to obtain $\mathbf{F}^{-1/2}$, is much less than the LMMSE estimator derived in [30] for quantized MIMO-OFDM. The estimator derived in [30] requires both a CP and the computation of the inverse of an $M\tau \times M\tau$ matrix. This is a large matrix as M is large in massive MIMO and $\tau \geq KL$ in general. Mean-square-error (MSE) matrix for $\hat{\underline{\mathbf{h}}}^{\text{LMMSE}}$ can also be calculated as

$$\begin{aligned}
\mathbf{C}_{\hat{\underline{\mathbf{h}}}}^{\text{LMMSE}} &= \mathbf{E}\big[\big(\hat{\underline{\mathbf{h}}}^{\text{LMMSE}} - \underline{\mathbf{h}}^{\text{LMMSE}}\big)\big(\hat{\underline{\mathbf{h}}}^{\text{LMMSE}} - \underline{\mathbf{h}}^{\text{LMMSE}}\big)^H\big] \\
&= \mathbf{I}_{MKL} - \left(\left(\mathbf{X}^H\mathbf{P}^H\mathbf{P}\mathbf{X} + \mathbf{I}_{KL}\right)^{-1} \otimes \mathbf{I}_M\right) \left(\mathbf{X}^H\mathbf{P}^H\mathbf{P}\mathbf{X} \otimes \mathbf{I}_M\right).
\end{aligned} \tag{7.17}$$

7.3.1 Low Complexity Approximations for the LMMSE Estimator

In this section, we will show that an even lower complexity approximation for LMMSE channel estimate exists under some conditions. One of those conditions is $\mathbf{X}\mathbf{X}^H$ being diagonally dominant. This happens when the pilots assigned

to different users are nearly orthogonal and the autocorrelation function of the transmitted pilot sequence of all users is close to an impulse function, that is, $\sum_{n=0}^{L-1}\sqrt{\rho_k[n]}\sqrt{\rho_{k'}[n]}x_k[m-n]x_{k'}^*[m'-n] \approx U\delta[m-m']\delta[k-k']$, where U is a multiplicative constant. This approximation is accurate when users transmit randomly generated complex symbols as pilot sequences and KL is large. Another condition for LMMSE channel estimate to have a lower complexity approximation is signal-to-noise ratio (SNR) being low. For both cases (for low SNR or when $\mathbf{X}\mathbf{X}^H$ is diagonally dominant) $\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}} = \left(\mathbf{X}\mathbf{X}^H \otimes \mathbf{I}_M\right) + N_o\mathbf{I}_{M\tau}$ is diagonally dominant, thus $\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}}$ can be approximated using (C.3)-(C.6) as a diagonal matrix as

$$\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}} \approx (2-4/\pi)\,\mathbf{I}_{M\tau}. \tag{7.18}$$

This means that the overall effective noise becomes uncorrelated also for the one-bit quantized case. Then, the complexity of the calculation of the whitening filter $\mathbf{C}_{\underline{\mathbf{\Gamma}}}^{-1/2} = \mathbf{F}^{-1/2} \otimes \mathbf{I}_M$ is reduced significantly as $\mathbf{F}$ is diagonal for uncorrelated effective noise. This makes $\mathbf{P} = \mathbf{F}^{-1/2}\mathbf{B}$ a diagonal matrix as $\mathbf{B}$ is a diagonal matrix, reducing the complexity in calculating (7.16).

A further reduction in the complexity of the LMMSE channel estimator is possible by assuming that $(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau)$ is a constant diagonal matrix. This is an accurate assumption when SNR is low. Even if SNR is not low, it is valid to assume that the diagonal elements of $\mathbf{X}\mathbf{X}^H$, which are equal to the average received power at each antenna, does not change over the pilot symbol transmission phase for most cases. If the magnitude of the transmitted complex pilot symbols are always the same, which is the case for DFT pilot sequences or any sequence generated randomly from a phase-shift keying (PSK) type modulation, this assumption is exactly correct. Otherwise, the approximation error caused by this assumption goes to zero as KL becomes large. With this assumption, $\mathbf{B}_1$ and $\mathbf{B}_m$ in (C.1) and (C.2) can be approximated as

$$\mathbf{B}_1 = \sqrt{4/\pi}\mathrm{diag}(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau)^{-0.5} \approx g_1\mathbf{I}_\tau, \quad \mathbf{B}_m \approx g_m\mathbf{I}_\tau, \tag{7.19}$$

where $g_1 = \sqrt{4/(\pi(P_r))}$, in which $P_r \triangleq KE_s + N_o$ is the average received power and

$$g_m = \sqrt{\Delta^2/\left(\pi P_r\right)} \times \sum_{i=1}^{2^q-1} \exp\left(-\Delta^2\left(i-2^{q-1}\right)^2/\sqrt{(P_r)}\right). \tag{7.20}$$

Along with (C.8) and (7.18), this implies that

$$\mathbf{P} = \mathbf{F}^{(-1/2)}\mathbf{B} \approx (d+g^2N_o)^{(-1/2)}g\mathbf{I}_\tau, \tag{7.21}$$

where $g = g_1\chi_q + g_m(1-\chi_q)$ and $d = (2-4/\pi)\chi_q + d_m(1-\chi_q)$, in which

$$\begin{aligned} d_m =& \frac{\Delta^2}{2}(2^q-1)^2 - g^2(P_r) \\ & - 4\Delta^2 \sum_{i=1}^{2^q-1} \left(i - 2^{q-1}\right) \\ & \times \left(1 - \mathcal{Q}\left(\sqrt{2}(i-2^{q-1})\,(P_r)^{-1/2}\right)\right). \end{aligned} \tag{7.22}$$

Then, the LMMSE estimator in (7.16) and MSE expression in (7.17) can be approximated as

$$\hat{\underline{\mathbf{h}}}^{\text{LMMSE}} \approx \left(\left(c^2\mathbf{X}^H\mathbf{X} + \mathbf{I}_{KL}\right)^{-1} \otimes \mathbf{I}_M\right)(c\mathbf{X}\otimes\mathbf{I}_M)^H \underline{\mathbf{z}}^{(p)}, \tag{7.23}$$

$$\mathbf{C}_{\hat{\underline{\mathbf{h}}}}^{\text{LMMSE}} \approx \mathbf{I}_{MKL} - \left(\left(c^2\mathbf{X}^H\mathbf{X} + \mathbf{I}_{KL}\right)^{-1} \otimes \mathbf{I}_M\right)\left(c^2\mathbf{X}^H\mathbf{X}\otimes\mathbf{I}_M\right), \tag{7.24}$$

where $c = g/\sqrt{d + g^2N_o}$. By approximating $\mathbf{X}\mathbf{X}^H$ as a constant diagonal matrix (with diagonal entries being $KE_s + N_o$), which is an accurate assumption under the conditions mentioned above, expressions of even lower complexity to calculate can be written from (7.14) as

$$\hat{\underline{\mathbf{h}}}^{\text{LMMSE}} \approx \frac{c\,(\mathbf{X}\otimes\mathbf{I}_M)^H\,\underline{\mathbf{z}}^{(p)}}{c^2E_sK+1}, \mathbf{C}_{\hat{\underline{\mathbf{h}}}}^{\text{LMMSE}} \approx \left(1 - \frac{c^2\left(\mathbf{X}^H\mathbf{X}\otimes\mathbf{I}_M\right)}{c^2E_sK+1}\right), \tag{7.25}$$

which are very simple expressions not involving any matrix inversions.

7.4 Data Transmission

For the data transmission phase, the quantized received signal can be rewritten using (7.5) as

$$\underline{\mathbf{r}}^{(d)} = \mathrm{Q}\left(\underline{\mathbf{y}}^{(d)}\right) = \mathrm{Q}\left(\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{w}}\right), \tag{7.26}$$

where

$$\begin{aligned} \underline{\mathbf{y}}^{(d)} &\triangleq \left[\mathbf{y}[0]^T\,\mathbf{y}[1]^T\cdots\mathbf{y}[N+L-2]^T\right]^T, \\ \underline{\mathbf{w}} &\triangleq \left[\mathbf{w}[0]^T\,\mathbf{w}[1]^T\cdots\mathbf{w}[N+L-2]^T\right]^T, \\ \underline{\mathbf{H}} &\triangleq \mathrm{blkToeplitz}(\underline{\mathbf{H}}^c, \underline{\mathbf{H}}^r), \\ \underline{\mathbf{x}} &\triangleq \left[\mathbf{x}[0]^T\,\mathbf{x}[1]^T\cdots\mathbf{x}[N-1]^T\right]^T, \end{aligned} \tag{7.27}$$

$$\begin{aligned} \underline{\mathbf{H}}^c &\triangleq \left[\mathbf{H}'[0]^T\,\mathbf{H}'[1]^T\,\cdots\,\mathbf{H}'[L-1]^T\,\mathbf{0}\cdots\mathbf{0}\right]^T, \\ \underline{\mathbf{H}}^r &\triangleq \left[\mathbf{H}'[0]\,\mathbf{0}\,\cdots\,\mathbf{0}\right], \mathbf{H}'[\ell] = \mathbf{H}[\ell]\mathbf{J}[\ell], \end{aligned} \tag{7.28}$$

in which N is the data packet length and the sizes of the matrices $\underline{\mathbf{H}}^c$ and $\underline{\mathbf{H}}^r$ are $(N+L-2)M \times K$ and $M \times NK$, respectively. Note that despite the same or similar notations are used for the data/pilot and noise vectors for the channel estimation and data transmission signal models for simplicity, they are completely independent of each other. Using Bussgang decomposition [32], (7.26) can be reexpressed as

$$\underline{\mathbf{r}}^{(d)} = \underline{\mathbf{A}}^{(d)}\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{A}}^{(d)}\underline{\mathbf{w}} + \underline{\mathbf{q}}^{(d)}, \tag{7.29}$$

where $\underline{\mathbf{A}}^{(d)}$ can be found by replacing $\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}$ in (C.1)-(C.2) with $\mathbf{C}_{\underline{\mathbf{y}}^{(d)}} \triangleq E_s\underline{\mathbf{H}}\underline{\mathbf{H}}^H + N_o\mathbf{I}_{M(N+L-2)}$. Moreover, the quantizer distortion term covariance matrix for one-bit quantized case, namely $\mathbf{C}^1_{\underline{\mathbf{q}}^{(d)}}$, can be found by replacing $\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}$ and $\underline{\mathbf{A}}^{(p,1)}$ in (C.3) and (C.4) with $\mathbf{C}_{\underline{\mathbf{y}}^{(d)}}$ and $\underline{\mathbf{A}}^{(d)}$, respectively. The quantizer distortion term autocovariance matrix for multi-bit quantized case, namely $\mathbf{C}^m_{\underline{\mathbf{q}}^{(d)}}$, can also be obtained by replacing $\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}$, $\underline{\mathbf{A}}^{(p,m)}$ and $\mathbf{I}_{M\tau}$ in (C.8) with $\mathbf{C}_{\underline{\mathbf{y}}^{(d)}}$, $\underline{\mathbf{A}}^{(d)}$ and $\mathbf{I}_{M(N+L-2)}$, respectively. Furthermore, when the channel coefficients are i.i.d. unit variance random variables, all diagonal elements of $\underline{\mathbf{H}}\underline{\mathbf{H}}^H$, which correspond to received average signal power, converge to KE_s when KL goes large. The reason for this convergence is the same as discussed previously for the pilot transmission phase. In addition, when KL is large or for low SNR, it is again straightforward to show that $\underline{\mathbf{H}}\underline{\mathbf{H}}^H + N_o\mathbf{I}_{M(N+L-2)}$ is a diagonally dominant matrix with diagonal entries converging to KE_s+N_o. Even for the spatially correlated channel case, diagonal elements will converge to $KE_s + N_o$ according to the Weak Law of Large Numbers as the channels observed by different users and different channel taps are still uncorrelated, although the diagonal dominance may be affected. In this case (when $\underline{\mathbf{H}}\underline{\mathbf{H}}^H + N_o\mathbf{I}_{M(N+L-2)} \approx (KE_s + N_o)\mathbf{I}_{M(N+L-2)}$), by employing the modified versions of (C.1), (C.3), (C.4), (C.2), (C.8) for data transmission phase (with the aforementioned modifications such as replacing $\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}$ by $\mathbf{C}_{\underline{\mathbf{y}}^{(d)}}$), $\underline{\mathbf{A}}^{(d)}$ and $\mathbf{C}_{\underline{\mathbf{q}}^{(d)}}$ can be approximated as

$$\underline{\mathbf{A}}^{(d)} \approx g\mathbf{I}_{M(N+L-2)}, \qquad \mathbf{C}_{\underline{\mathbf{q}}^{(d)}} \approx d\mathbf{I}_{M(N+L-2)}. \tag{7.30}$$

The aforementioned assumptions, implying an uncorrelated quantizer noise assumption, are observed to be accurate even when the SNR is high and the number of users are as low as $K = 4$ and $L = 1$ for i.i.d. channel coefficients [30, Fig. 4]. In fact, K and L values will be much larger in general, resulting in very low approximation errors.

Based on (7.29) and (7.30), a minimum distance performance metric can be constructed as

$$\Lambda\left(\underline{\underline{\mathbf{r}}}, \underline{\underline{\hat{\mathbf{H}}}}, \underline{\mathbf{x}}\right) = \gamma_1 \exp\left(-||\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\mathbf{H}}}}\,\underline{\mathbf{x}}||^2\right), \tag{7.31}$$

where γ_1 is a multiplicative constant, $\underline{\underline{\mathbf{r}}} = \underline{\mathbf{r}}/\sqrt{d+g^2N_o}$, $\underline{\underline{\hat{\mathbf{H}}}} = g\underline{\hat{\mathbf{H}}}/\sqrt{d+g^2N_o}$. Here $\underline{\hat{\mathbf{H}}}$ is the estimated version of the channel matrix $\underline{\mathbf{H}}$, which is found by using (7.27) and (7.28) such that $\mathbf{H}[\ell]$ in (7.28) is replaced by its estimate $\hat{\mathbf{H}}[\ell]$. Matrix $\hat{\mathbf{H}}[\ell]$ can be constructed by replacing the elements of $\mathbf{H}[\ell]$, which is defined in (7.5), with the corresponding the LMMSE channel coefficient estimates obtained with (7.16), (7.23) or (7.25) in Section 7.3. The metric in (7.31) corresponds to the ML metric when the effective noise term $(g\underline{\mathbf{w}}) + \underline{\mathbf{q}}^{(\mathbf{d})}$ has a Gaussian distribution. It has been pointed out in [45,96,97] that the Gaussian assumption for the effective noise $(g\underline{\mathbf{w}}) + \underline{\mathbf{q}}^{(\mathbf{d})}$ yields accurate results, especially for low SNR, even for 1-bit quantizer. With this finding, it can be stated that the effective noise $(g\underline{\mathbf{w}}) + \underline{\mathbf{q}}^{(\mathbf{d})}$ can also be approximated as Gaussian for higher quantization resolutions. The reason is that the $(g\underline{\mathbf{w}})$ term in the effective noise dominates for higher quantizer resolution as g gets closer to 1 and the power of quantizer noise $\underline{\mathbf{q}}^{(\mathbf{d})}$ decreases. Therefore, there are many studies that approximates the quantization noise as Gaussian [45, 96–101].

We continue by rewriting the minimum distance metric in (7.31) as

$$\Lambda\left(\underline{\underline{\mathbf{r}}}, \underline{\underline{\hat{\mathbf{H}}}}, \underline{\mathbf{x}}\right) = \gamma_2 \exp\left(2\,\mathrm{Re}\left(\underline{\underline{\mathbf{r}}}^H \underline{\underline{\hat{\mathbf{H}}}}\,\underline{\mathbf{x}}\right) - \underline{\mathbf{x}}^H \underline{\underline{\hat{\mathbf{H}}}}^H \underline{\underline{\hat{\mathbf{H}}}}\,\underline{\mathbf{x}}\right). \tag{7.32}$$

To obtain the optimal estimates based on (7.32), there are various approaches. One is to filter $\underline{\underline{\mathbf{r}}}$ by a channel matched filter (CMF) followed by a noise whitening filter in the Forney method [102]. The complexity of this method can be high due to whitening filter, thus an alternative method based on Ungerboeck observation model can be adopted [102]. In the Ungerboeck observation model, the minimum distance metric is constructed directly from the unwhitened CMF output, namely $\mathbf{v} \triangleq \underline{\underline{\hat{\mathbf{H}}}}^H \underline{\underline{\mathbf{r}}}$. Taking $\mathbf{v}$ as the observation vector, the metric in (7.32) can be rewritten as

$$\Lambda\left(\underline{\underline{\mathbf{r}}}, \underline{\underline{\hat{\mathbf{H}}}}, \underline{\mathbf{x}}\right) = \gamma_2 \exp\left(2\,\mathrm{Re}\left(\mathbf{v}^H \underline{\mathbf{x}}\right) - \underline{\mathbf{x}}^H \mathbf{G}\,\underline{\mathbf{x}}\right), \tag{7.33}$$

where

$$\mathbf{G} \triangleq \underline{\underline{\hat{\mathbf{H}}}}^H \underline{\underline{\hat{\mathbf{H}}}} \triangleq \mathrm{blkToeplitz}(\underline{\mathbf{G}}^c, \underline{\mathbf{G}}^r),$$

in which

$$\begin{aligned}
\mathbf{G}^r &= [\mathbf{G}[0]\ \mathbf{G}[1]\ \cdots\ \mathbf{G}[L-1]\ \mathbf{0}\ \cdots\ \mathbf{0}]\,,\\
\mathbf{G}^c &= (\mathbf{G}^r)^H\,,\\
\mathbf{G}[\ell] &\triangleq g^2/(d+g^2N_o)\sum_{k=0}^{L-1-\ell}\hat{\mathbf{H}}'[k+\ell]^H\hat{\mathbf{H}}'[k],\\
\hat{\mathbf{H}}'[\ell] &\triangleq \hat{\mathbf{H}}[\ell]\mathbf{J}[\ell].
\end{aligned}$$

With these definitions, the minimum distance metric in (7.33) can be computed recursively as

$$\begin{aligned}
\ln(\Lambda(.)) = \sum_{n=0}^{N-1}\Bigg(\sum_{k=1}^{K}\Bigg[&\kappa_k^n(v_k[n],x_k[n]) - \phi_k^n(x_k[n],\mathbf{s}_k^n)\\
&- \sum_{k'=1,k'<k}^{K}\psi_{k,k'}^n(x_k[n],\mathbf{s}_k^n,x_{k'}[n],\mathbf{s}_{k'}^n)\Bigg]\Bigg)\,,
\end{aligned} \tag{7.34}$$

where $\Lambda(.) = \Lambda\left(\mathbf{r},\underline{\underline{\hat{\mathbf{H}}}},\underline{\mathbf{x}}\right)$, $v_k[n]$ is the $(Kn+k)^{th}$ element of $\mathbf{v}$, and $\mathbf{s}_k^n$ is the state vector of user k at the nth time instant, which can be expressed as

$$\mathbf{s}_k^n = [x_k[n-1]\ \cdots\ x_k[n-J]]\,. \tag{7.35}$$

As can be noted in (7.35), although we need to have $L-1$ elements in $\mathbf{s}_k^n$ for optimal sequence estimation, the number of elements in the state vector in (7.35), namely J, can be selected to be less than $L-1$, to reduce the complexity of the detector. This can be done by constructing surviving paths based on the proposed Ungerboeck-type reduced state sequence estimation (U-RSSE) with bidirectional decision feedback algorithm for MIMO, the details of which will be provided in the sequel. The functions $\kappa_k^n(.)$, ϕ_k^n and $\psi_{k,k'}^n(.)$ in (7.34) are also defined as

$$\kappa_k^n(.) \triangleq 2\,\mathrm{Re}\left\{(v_k^*[n])x_k[n]\right\} - x_k^*[n][\mathbf{G}[0]]_{(k,k)}x_k[n], \tag{7.36}$$

$$\phi_k^n(.) \triangleq 2\,\mathrm{Re}\left\{\zeta_{k,k}[n]\right\}, \psi_{k,k'}^n(.) \triangleq 2\,\mathrm{Re}\left\{x_{k'}^*[n][\mathbf{G}[0]]_{(k',k)}x_k[n] + \zeta_{k,k'}[n] + \zeta_{k',k}[n]\right\}, \tag{7.37}$$

where $\zeta_{k,k'}[n] = \sum_{\ell=1}^{\min(L-1,n)} x_k^*[n][\mathbf{G}[\ell]]_{(k,k')}^H x_{k'}[n-\ell]$. Here, $\kappa_k^n(.)$ can be regarded as the CMF output, $\phi_k^n(.)$ calculates the self-interference due to ISI, while $\psi_{k,k'}^n(.)$ corresponds to the interference caused by the other users to user k. The metric in

(7.34) can be redefined taking into account the *a priori* probabilities of the transmitted data symbols as

$$\ln\Bigg(\Lambda\Bigg(\{x_k[n],\ \mathbf{s}_k^n, v_k^n\}_{\forall k,n}\Bigg)\Bigg)$$
$$\propto \sum_{n=0}^{N-1}\sum_{k=1}^{K}\Bigg\{\ln\left(\Pr\left(\mathbf{s}_k^0\right)\right) + \kappa_k^n\left(v_k[n], x_k[n]\right) - \phi_k^n\left(x_k[n], \mathbf{s}_k^n\right)$$
$$+ \ln\left(T_k^n\left(x_k[n], \mathbf{s}_k^n, \mathbf{s}_k^{n+1}\right)\right) + \ln\left(\Pr\left(\{x_k[n]\}\right)\right)$$
$$- \sum_{k'=1,k'<k}^{K} \psi_{k,k'}^n\left(x_k[n], \mathbf{s}_k^n, x_{k'}[n], \mathbf{s}_{k'}^n\right)\Bigg\}, \tag{7.38}$$

where $\Pr(\{x_k[n]\})$ and $\Pr\left(\mathbf{s}_k^0\right)$ are the *a priori* probabilities of the data symbol $x_k[n]$ and the initial state vector $\mathbf{s}_k^0$. Moreover, $\ln(.)$ takes the natural logarithm, and $T_k^n\left(x_k[n], \mathbf{s}_k^n, \mathbf{s}_k^{n+1}\right)$ is the trellis indicator function, which is equal to 1 if a transition from $\mathbf{s}_k^n$ to $\mathbf{s}_k^{n+1}$ is possible with the data symbol being $x_k[n]$. Otherwise, it is equal to zero. The proposed factor graph (FG) constructed for the calculation of (7.38) is presented in Fig. 7.1. As can be noted in Fig. 7.1, there are cycles of length 6. Although the existence of cycles in the FG in Fig. 7.1 result in approximate computation of *a posteriori* probabilities (APP) of each transmitted symbol, the approximation errors due to cycles are known to be negligable if the length of the cycles are greater than 4 [103]. As can also be noted in Fig. 7.1, the state vector $\mathbf{s}_k^n$ and the data symbol $x_k[n]$ are merged into a single variable node in order to increase the cycle length, which is known as streching in the literature [104].

Based on the FG in Fig. 7.1, a novel reduced complexity quantization-aware Ungerboeck-type message passing algorithm with bidirectional decision feedback (QA-UMPA-BDF) detector is proposed. The proposed detector is characterized by the following message update rules based on the sum-product algorithm (SPA) framework:

$$\Lambda_k^{f,n+1}\left(\mathbf{s}_k^{n+1}\right) = \ln\Bigg(\sum\nolimits_{\sim\{\mathbf{s}_k^{n+1}\}} \exp\big(\Lambda_k^{f,n}\left(\mathbf{s}_k^n\right) + \ln(T_k^n\left(.\right))$$
$$- \phi_k^n\left(.\right) + V_k^n\left(x_k[n], \mathbf{s}_k^n\right)\big)\Bigg), \tag{7.39}$$

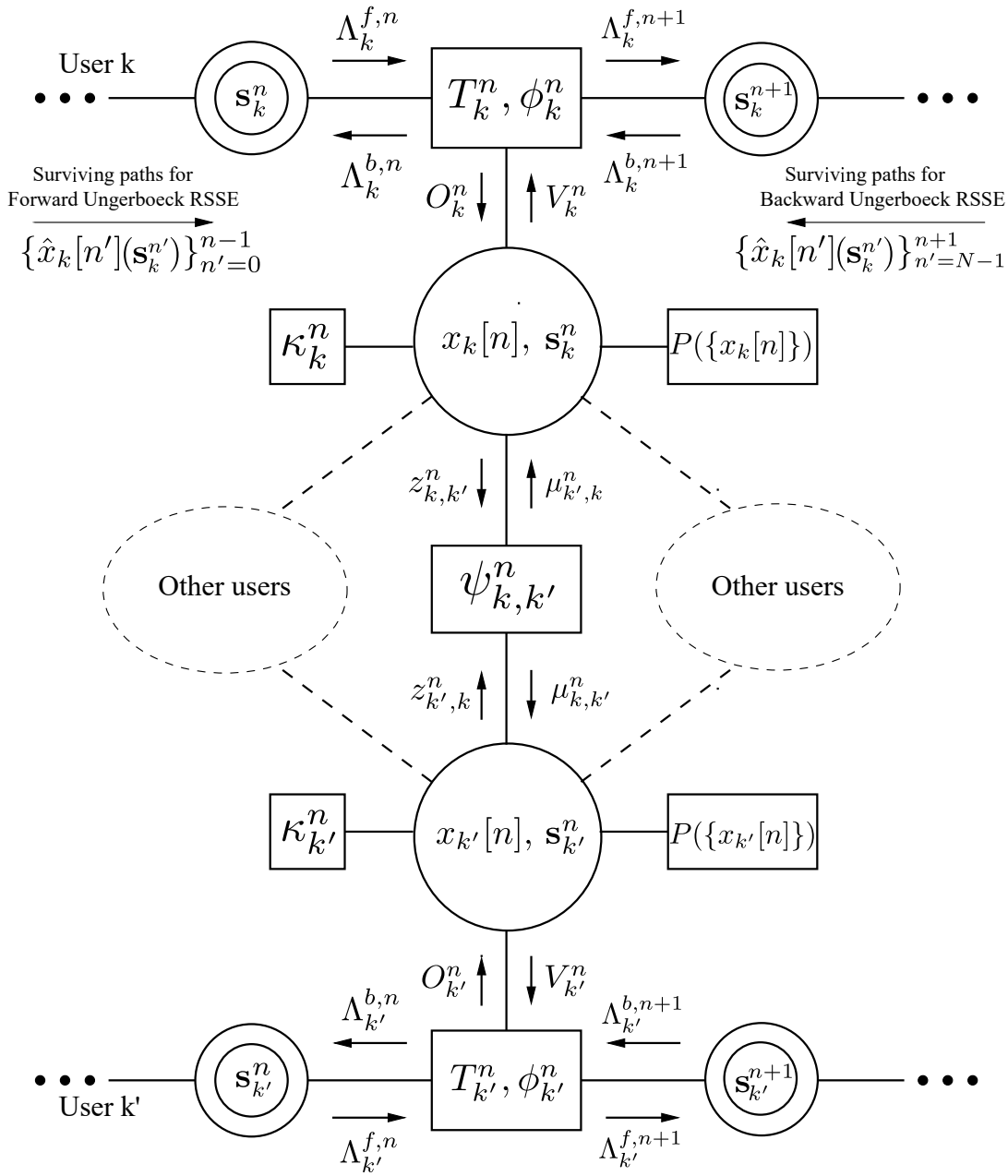


Figure 7.1: Proposed factor graph corresponding to the calculation of the metric in (7.38).

$$\Lambda_k^{b,n}\left(\mathbf{s}_k^n\right) = \ln\Bigg(\sum\nolimits_{\sim\{\mathbf{s}_k^n\}} \exp\big(\Lambda_k^{b,n+1}\left(\mathbf{s}_k^{n+1}\right) + \ln\left(T_k^n\left(.\right)\right) - \phi_k^n\left(.\right) + V_k^n\left(x_k[n], \mathbf{s}_k^n\right)\big)\Bigg), \tag{7.40}$$

$$O_k^n\left(x_k[n], \mathbf{s}_k^n\right) = \Lambda_k^{f,n}\left(\mathbf{s}_k^n\right) + \Lambda_k^{b,n+1}\left(\mathbf{s}_k^{n+1}\right) + \ln\left(T_k^n\left(.\right)\right) - \phi_k^n\left(.\right), \tag{7.41}$$

$$V_k^n\left(x_k[n], \mathbf{s}_k^n\right) = \ln\left(\Pr\left(\{x_k[n]\}\right)\right) + \kappa_k^n\left(.\right) + \sum_{\{l=1,\, l\neq k\}}^{K} \mu_{l,k}^n\left(x_k[n], \mathbf{s}_k^n\right), \tag{7.42}$$

$$\mu_{k',k}^n\left(x_k[n], \mathbf{s}_k^n\right) = \ln\Bigg(\sum\nolimits_{\{x_{k'}[n], \mathbf{s}_{k'}^n\}} \exp\left(z_{k',k}^n\left(x_{k'}[n], \mathbf{s}_{k'}^n\right) - \psi_{k,k'}^n\left(.\right)\right)\Bigg), \tag{7.43}$$

$$z_{k,k'}^n\left(x_k[n], \mathbf{s}_k^n\right) = O_k^n\left(x_k[n], \mathbf{s}_k^n\right) + V_k^n\left(x_k[n], \mathbf{s}_k^n\right) - \mu_{k',k}^n\left(x_k[n], \mathbf{s}_k^n\right), \tag{7.44}$$

where $\sum_{\sim\{x\}}$ is defined as the sum over all variables excluding x. Note that we calculate messages in log-domain to avoid numerical issues stemming from large numbers as multiplications performed in SPA are reflected as summations in log domain in (7.39)-(7.44). For further avoidance of numerical issues, the max-log approximation [105] is used for (7.39), (7.40) and (7.43) as

$$\Lambda_k^{f,n+1}\left(\mathbf{s}_k^{n+1}\right) \approx \max_{\{\mathbf{s}_k^n\}} \left(\Lambda_k^{f,n}\left(\mathbf{s}_k^n\right) + \ln(T_k^n\left(.\right)) - \phi_k^n\left(.\right) + V_k^n\left(x_k[n], \mathbf{s}_k^n\right)\right), \tag{7.45}$$

$$\Lambda_k^{b,n}\left(\mathbf{s}_k^n\right) \approx \max_{\{\mathbf{s}_k^{n+1}\}} \left(\Lambda_k^{b,n+1}\left(\mathbf{s}_k^{n+1}\right) + \ln(T_k^n\left(.\right)) - \phi_k^n\left(.\right) + V_k^n\left(x_k[n], \mathbf{s}_k^n\right)\right), \tag{7.46}$$

$$\mu_{k',k}^n\left(x_k[n], \mathbf{s}_k^n\right) \approx \max_{\{x_{k'}[n], \mathbf{s}_{k'}^n\}} \left(z_{k',k}^n\left(x_{k'}[n], \mathbf{s}_{k'}^n\right) - \psi_{k,k'}^n\left(.\right)\right). \tag{7.47}$$

7.4.1 Bias Compensation

Owing to the state reduction and the pre-cursor ISI that remains after CMF operation, an anti-causal interference appears. As a result, U-RSSE suffers from *correct path loss* even when there is no noise and multi-user interference, as pointed out for unquantized single-input single-output (SISO) systems [106]. This interference results in a *bias* affecting the tentative decisions in a survivor map. This bias has to be corrected in the forward surviving path construction. With such a correction, the surviving path for the states of the kth user can be constructed as

$$\hat{x}_k[n-J](\mathbf{s}_k^n) = \arg\max_{x_k[n-J]} \left[\Lambda_k^{f,n}\left(S_k^n\right) + \phi_k^n\left(.\right) + V_k^n\left(.\right) - \beta_k^{n-J}\left(\mathbf{s}_k^n, x_k[n-J]\right)\right], \tag{7.48}$$

where $\beta_k^{n-J}(.)$ is the *bias* correction term. The bias correction term can be calculated by extending the bias derivation made in [106] for SISO case to multi-user case as

$$\beta_k^{n-J}\left(\mathbf{s}_k^n, x_k[n-J]\right) = 2\,\mathrm{Re}\left\{\sum_{k'=1}^{K}\sum_{l_1=n-L+2}^{n-J}\sum_{l_2=n-l_1+1}^{L-1}\Big[x_k^*[l_1](\mathbf{s}_k^n)[\mathbf{G}[l_2]]_{(k,k')} \times \tilde{x}_{k'}[l_1+l_2]\Big]\right\}, \quad (7.49)$$

where $x_k^*[l_1](\mathbf{s}_k^n)$ for $l_1 < n-j$ can be found from the surviving paths constructed using (7.48) at the previous time instants. Moreover, $\tilde{x}_{k'}[l_1+l_2]$ can also be found from the hard tentative decisions about future symbols, obtained in the previous iterations (what is meant by "iterations" will be detailed in Section 7.4.2). The bias term is also simplified for the full decision feedback case (when no state is used, that is, when $J=0$) as

$$\beta_k^n\left(x_k[n]\right) = 2\,\mathrm{Re}\left\{\sum_{k'=1}^{K}\left[\sum_{l_2=1}^{L-1} x_k^*[n][\mathbf{G}[l_2]]_{(k,k')}\tilde{x}_{k'}[n+l_2]\right]\right\}, \quad (7.50)$$

since the terms of the outer summation with index $\ell_1 \neq n-J = n$ in (7.49) can be omitted as the maximization is over $x_k[n]$ in (7.48) for $J=0$.

The marginalized version of the metric in (7.38) can be calculated in the termination step as

$$\ln\left(\Lambda\left(x_k[n], \mathbf{s}_k^n, \{v_k^n\}_{\forall k,n}\right)\right) = \sum_{S_k^n}\left[V_k^n(.) + O_k^n(.) - \beta_k^n\left(\mathbf{s}_k^n, x_k[n]\right)\right], \quad (7.51)$$

with the corresponding data symbol estimates maximizing the metric in (7.38) given as

$$\hat{x}_k[n] = \underset{\{x_k[n]\}}{\arg\max}\sum_{\mathbf{s}_k^n}\left[V_k^n\left(x_k[n], \mathbf{s}_k^n\right) + O_k^n\left(x_k[n], \mathbf{s}_k^n\right) - \beta_k^n\left(\mathbf{s}_k^n, x_k[n]\right)\right]. \quad (7.52)$$

7.4.2 Message Passing Schedule

Owing to the cycles existing in the FG in Fig. 7.1, there is no unique message passing schedule for SPA operation. Therefore, we employ a serial schedule as in [104] for updating the messages. The proposed scheduling for forward recursion in time-domain is presented in Algorithm 1.

Algorithm 1 QA-UMPA-BDF, forward recursion in time-domain

Input: MF output $\mathbf{v}$ and correlation metric $\mathbf{G}$.

Initialization: Initialize all messages $z^n_{k',k}$, $\mu^n_{k',k}$, V^n_k, O^n_k, $\Lambda^{f,n}_k$, $\Lambda^{b,n}_k$ as zero.

for $n = 0 : 1 : N-1$ **do**
  for $k = 1 : 1 : K$ **do**
    `Update` $O^n_k(.)$ `using (7.41).`
  end for
  Forward recursion in user domain:
  for $k = 1 : 1 : K$ **do**
    for $k' = 1 : 1 : k-1$ **do**
      `Update` $\mu^n_{k',k}$ `using (7.47).`
    end for
    `Update the term` $V^n_k(.)$ `using (7.42).`
    for $k' = k+1 : 1 : K$ **do**
      `Update` $z^n_{k,k'}$ `using (7.44).`
    end for
  end for
  Backward recursion in user domain:
  for $k = K : -1 : 1$ **do**
    for $k' = K : -1 : k+1$ **do**
      `Update` $\mu^n_{k',k}$ `using (7.47).`
    end for
    `Update the term` $V^n_k(.)$ `using (7.42).`
    for $k' = k-1 : -1 : 1$ **do**
      `Update` $z^n_{k,k'}$ `using (7.44).`
    end for
  end for
  Update the time-domain forward messages:
  for $k = 1 : 1 : K$ **do**
    `Update` $\Lambda^{f,n+1}_k(.)$ `using (7.45).`
    `Calculate bias term` $\beta^{n-J}_k(\mathbf{s}^n_k, x_k[n-J])$ `using (7.49) or (7.50).`
    `Update surviving paths` $\hat{x}_k[n-J](\mathbf{s}^n_k)$ `using (7.48).`
  end for
end for

When the forward recursion in time-domain in Algorithm 1 ends, the same procedure is performed as the backward recursion in time-domain, except that the time index at the outermost for-loop in Algorithm 1 will be from $N-1$ to 0, the operation in line 27 will be replaced by an update of $\Lambda_k^{b,n}$ using (7.46), and the lines 28-29 will not be performed. Completion of forward and backward recursions in time-domain constitutes an iteration of QA-UMPA-BDF. Although various choices can be made for stopping criteria, the one adopted in this study is the completion of a predefined number of iterations. The initialization step in Algorithm 1 should only be performed for the forward recursion in time-domain at the first iteration.

Algorithm 1 has all details about how the messages in the FG in Fig. 7.1, which depicts the fundamental structure of the QA-UMPA-BDF algorithm, will be updated according to which schedule and equations among (7.39)-(7.47). Moreover, Algorithm 1 also specifies at which point of QA-UMPA-BDF algorithm, bias compensation and surviving path construction will be made according to which equations among (7.48)-(7.50). Therefore, it can be seen as the main description of an iteration of the QA-UMPA-BDF algorithm. After the necessary number of iterations of QA-UMPA-BDF algorithm is reached, the final data symbol estimates are found using (7.51)-(7.52).

7.4.3 Computational Complexity Analysis

The computational complexity per iteration of the proposed QA-UMPA-BDF detector can be found by analyzing (7.39)-(7.49). For the complexity analysis we consider the max-log approximations for (7.39), (7.40) and (7.43), which are (7.45), (7.46), (7.47). The complexity (number of flops) to calculate the messages per single iteration of the proposed detector is provided in Table 7.2, in which P is the modulation size.

Table 7.2: Computational complexity of the QA-UMPA-BDF detector per iteration.

	(7.41)	(7.42)-(7.44)	(7.45), (7.46)	(7.47)	(7.48), (7.49)
Complexity	$\mathcal{O}\left(NP^{(J+1)}KL\right)$	$\mathcal{O}\left(NP^{(J+1)}K^2L\right)$	$\mathcal{O}\left(NP^{(J+1)}K\right)$	$\mathcal{O}\left(NP^{2(J+1)}K\right)$	$\mathcal{O}\left(NP^{(J+1)}K^2L\right)$

As can be noted, the computational complexity per iteration can be as high as $\mathcal{O}\left(NP^{2(L+1)}K\right)$ if reduced state estimation is not employed (when $J=L-1$). How-

ever, the computational complexity can be reduced to $\mathcal{O}(NPK^2) + \mathcal{O}(NP^2K) + \mathcal{O}(NPK^2L)$ for $J = 0$, which changes linearly with N, L, and quadratically with K and P. The complexity to calculate CMF output $\mathbf{v}$ and the correlation metric $\mathbf{G}$ are $\mathcal{O}(NMKL)$ and $\mathcal{O}(MK^2L)$. Therefore, the total complexity of the QA-UMPA-BDF detector is $\mathcal{O}(INP^2K)+\mathcal{O}(INPK^2L)+\mathcal{O}(NMKL)+\mathcal{O}(MK^2L)$, where I is the number of iterations. The computational complexity of the representative benchmark algorithm that we compare the proposed QA-UMPA-BDF detector, namely the "Robust MMSE" in [86, Eqn.(27)], is $\mathcal{O}(MKN\log_2(N))+\mathcal{O}(NMK)+\mathcal{O}(NMK^2)+\mathcal{O}(NK^3)+\mathcal{O}(NKP)$, which grows with K^3. Therefore, the proposed QA-UMPA-BDF detector for $J = 0$ has lower complexity compared to the benchmark detector, especially when K is large. We will also show in Section 7.5 that the proposed detector can converge in about $I = 2$ iterations for most of the cases. Therefore, the number of iterations does not increase the proposed detector complexity to a significant degree. Detectors other than the "Robust MMSE" detector are also proposed in [86]. However, their performance is considered to be inferior compared to "Robust MMSE" detector [86, Fig.12,13], thus "Robust MMSE" is chosen as the benchmark detector.

7.5 Performance Metrics and Simulation Results

To assess the performance of the proposed LMMSE channel estimator, normalized MSE (nMSE) will be used as a metric. The nMSE taking into account the channel coefficients multiplied by the PDP can be found as [107]

$$\text{nMSE} = \frac{\text{Tr}\left[\mathbf{\Omega}\mathbf{C}_{\underline{\hat{\mathbf{h}}}}^{\text{LMMSE}}\mathbf{\Omega}^H\right]}{\text{Tr}\left[\mathbf{\Omega}\mathbf{C}_{\underline{\mathbf{h}}}\mathbf{\Omega}^H\right]} = \frac{\text{Tr}\left[\mathbf{\Omega}\mathbf{C}_{\underline{\hat{\mathbf{h}}}}^{\text{LMMSE}}\mathbf{\Omega}^H\right]}{MK}, \tag{7.53}$$

where $\mathbf{\Omega}$ is a diagonal matrix whose $(ML(k-1)+M\ell+1)^{th}$ to $(ML(k-1)+M\ell+M)^{th}$ diagonal elements are all equal to $\sqrt{\rho_k[\ell]}$. $\mathbf{C}_{\underline{\hat{\mathbf{h}}}}^{\text{LMMSE}}$ can be found from (7.17), (7.24) or (7.25). For the data detector performance metric, we use uncoded bit-error-rate (BER) and average mismatched achievable rate (AIR) per user [108]. Average mismatched AIR is a suitable metric to assess the performance of mismatched detectors employing approximate APPs for detection, as the exact APPs cannot be calculated due to the cycles in the FG in Fig. 7.1 and Gaussian effective noise ap-

proximations. The mismatched average AIR per user can be expressed as [108]

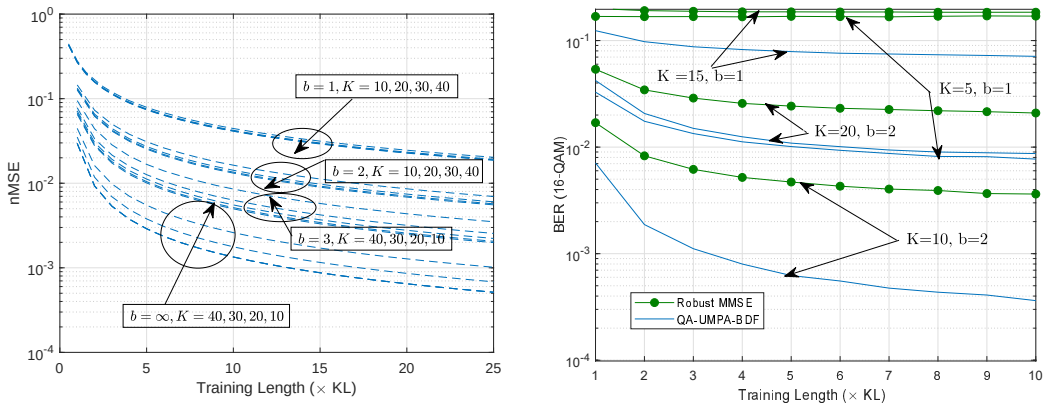
$$\mathrm{AIR} = \mathbf{E}_{\underline{\mathbf{x}},\underline{\mathbf{H}}}\left[\frac{1}{NK}\sum_{k=1}^{K}\sum_{n=0}^{N-1}\left[\log_2(P) - \log_2\left(\frac{\sum_{x'_k[n]\in A_x}\tilde{p}(\mathbf{v}|x'_k[n])}{\tilde{p}(\mathbf{v}|\hat{x}_k[n]=x_k[n])}\right)\right]\right], \quad (7.54)$$

where A_x is the set of all possible constellation points, $x_k[n]$ is the correct value of the transmitted symbol, and $\tilde{p}(\mathbf{v}|x_k[n])$ are the approximate APPs which are found using (7.52) as

$$\tilde{p}(\mathbf{v}|x_k[n]) \propto \sum_{\mathbf{s}_k^n}\exp\left(V_k^n\left(x_k[n],\mathbf{s}_k^n\right) + O_k^n\left(x_k[n],\mathbf{s}_k^n\right) - \beta_k^n\left(x_k[n],\mathbf{s}_k^n\right)\right). \quad (7.55)$$

Throughout the simulations, we will mostly concentrate on the performance comparison between the proposed QA-UMPA-BDF detector and the representative robust MMSE detector [86] from the literature. As the reduced state length of the QA-UMPA-BDF detector is set as $J = 0$, the representative detector has a comparable complexity to QA-UMPA-BDF detector. We will see that the proposed detector outperforms the representative detector in all cases, even if their complexities are similar and the QA-UMPA-BDF detector provides a higher spectral efficiency, due to the absence of a cyclic-prefix. Unless otherwise stated, $M = 100$, the PDP of the transmission channel is COST-207 typical delay profile for suburban and urban areas [109]. The number of channel taps $L = 32$, with the power ratio of the first and the last taps being 30 dB and $N = 1024$. The number of iterations for the QA-UMPA-BDF detector is selected as $I = 2$. The pilot symbols are created as random complex numbers from QPSK modulation. $E_b \triangleq E_s/\log_2(P)$ corresponds to the bit-energy. LMMSE channel estimates and MSE values are found based on (7.23) and (7.24). The step size of the quantizer is selected to optimally to minimize quantization noise as in [78]. The channel type is uncorrelated Rayleigh fading channel in all simulations except Fig. 7.4. For correlated channel cases, the antenna array type is assumed to be ULA, the mean arriving angles ϕ_k^ℓ are chosen from a uniform distribution between -45 and 45 degrees, and the angular spread is taken as $\varsigma_k^\ell = 2$ degrees. For Rician fading case (Fig. 7.4b), the Rician factor $\kappa_k[\ell] = 10$ dB, which is the case corresponding to an average of 100 meters distance to the base station from the user terminals and the scattering clusters [93]. Moreover, for Rician fading $L = 5$ instead of $L = 32$, as spatial correlation with Rician fading is mostly observed in mmWave scenarios, for which the number of taps is small in general.

To determine the necessary training length for the channel estimation, the nMSE or BER vs. the training length (τ) performances are obtained as in Fig. 7.2. In Fig. 7.2a,



(a) nMSE vs. training length (τ), $E_b/N_o = 0$ dB, $K = 10, 20, 30, 40$.

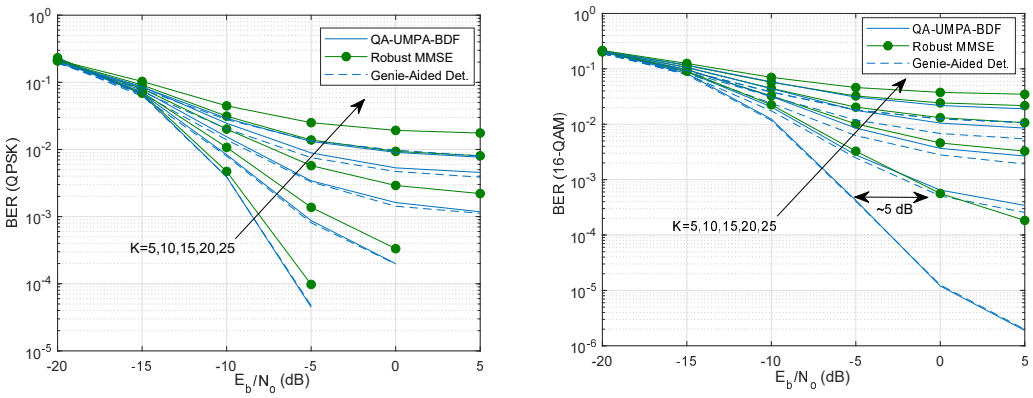
(b) BER vs. training length, $K = 5, 10, 15, 20$, $E_b/N_o = 0$ dB, 16-QAM.

Figure 7.2: nMSE (a) and BER (b) vs. training length (τ).

it seems that we have lower nMSE values when the number of users are increased. This is owing to the fact that the horizontal axis is provided as a multiple of KL, implying that the training length is higher for the same value of the horizontal axis if the number of users is higher. Since longer training length implies a higher received pilot energy and the pilots of users are nearly orthogonal, this results in lower nMSE for the channel estimates. We also observe in Fig. 7.2a that if we need to have a nMSE level less than 10^{-1}, $\tau \geq 5KL$ will be an adequate choice for the proposed LMMSE channel estimator for one-bit quantizer, although this number is much less for higher bit resolutions. However, we can also say that there is a decreased improvement for nMSE if the training length $\tau > 5KL$ for any bit resolution. Nevertheless, observing nMSE alone may not be enough to foresee how the error-rate performance of the proposed detector changes with the training length. Therefore, BER vs. training length is also obtained for 16-QAM modulated data symbols as in Fig. 7.2b. As can be noted in Fig. 7.2b, for 1 and 2 bits, there is a significant error-floor advantage of the proposed detector compared to the Robust MMSE detector [86] for all cases. Moreover, we can see that with very low resolution quantizers (1 or 2 bits), increasing τ more than $5KL$ is not very effective for decreasing BER. Therefore, we will set $\tau = 5KL$ when $q = 1, 2$. For $q = 3$, we will set $\tau = 3KL$ as the corresponding nMSE values close to 10^{-2} observed in Fig. 7.2, are considered to be adequate. For

$q > 3$, τ will be selected as $2KL$ in the subsequent simulations, all performed under imperfect channel state information (CSI).

In Fig. 7.3, we compare the BER performance of the proposed QA-UMPA-BDF and the Robust MMSE [86] detectors for either QPSK with $q = 1$ or 16-QAM with $q = 2$. We also include the performance of a genie aided detector, which is referred to as "Genie Aided Det." in all figures. This detector calculates the metric in (7.38) for an $x_k[n]$ assuming that all other symbols are perfectly known so that ISI and MUI terms in (7.38) are also calculated perfectly[1]. Genie aided detector performance is mainly limited by thermal and quantization noise.



(a) BER vs. E_b/N_o for QPSK, $q = 1$.

(b) BER vs. E_b/N_o for 16-QAM, $q = 2$.

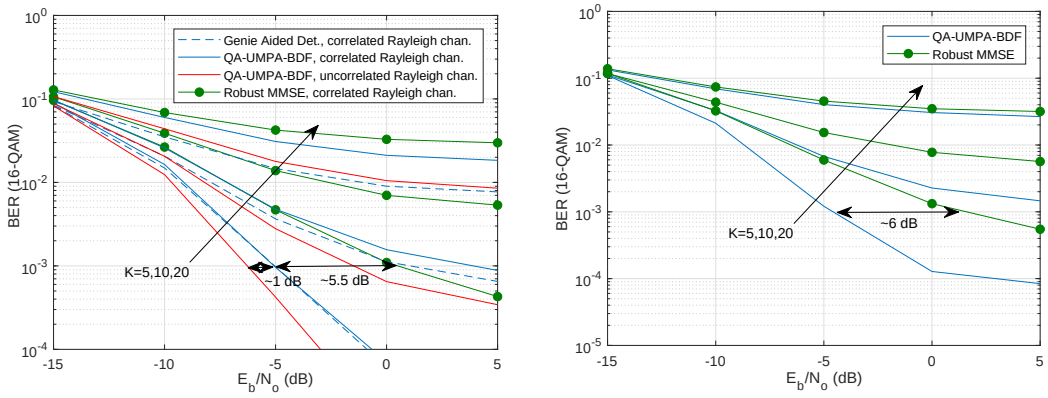
Figure 7.3: BER vs. E_b/N_o for QPSK, $q = 1$ (a) or 16-QAM, $q = 2$ (b)

As can be observed in Fig. 7.3a, the QA-UMPA-BDF detector has better performance compared to the representative benchmark detector for all number of user values (for $K = 5, 10, \ldots, 25$), despite being spectrally more efficient due to the CP free transmission. If the modulation type is changed to 16-QAM, the SNR advantage of QA-UMPA-BDF is up to 5 dB as can be noted in Fig. 7.3b. Moreover, the QA-UMPA-BDF performance is always very close to genie-aided detector performance with only 2 iterations, which is the case in most of the subsequent simulations.

We also investigate the error-rate performance of the QA-UMPA-BDF detector under spatially correlated channel in Fig. 7.4. For spatially correlated Rayleigh fading channel in Fig. 7.4a, the SNR advantage over the benchmark detector is observed as

[1] Genie aided detector uses perfect bidirectional decision feedback while constructing surviving paths in (7.48) and bias terms in (7.49). For unquantized case and perfect CSI, the performance of this detector corresponds to matched filter bound [106].

even more (5.5 dB vs. 5 dB for uncorrelated Rayleigh fading case). This can be attributed to the noise amplification effect due to an ill-conditioned matrix inversion in the benchmark receiver for spatially correlated case. In addition, despite the severe losses in diversity and multi-user interference suppression capability of the antenna array due to channel correlation, we see that the performance loss compared to uncorrelated channel case is less than 1 dB for $K = 5$. This verifies that the errors in the approximations in Section 7.3-7.4 relying on the diagonal dominance of $\mathbf{H}\mathbf{H}^H$ and diagonal $\mathbf{C}_{\underline{\mathbf{q}}^{(p)}}$ are limited for the correlated channel case. Moreover, we also present BER performance of QA-UMPA-BDF for correlated Rician fading in Fig. 7.4b. As can be noted, QA-UMPA-BDF outperforms the benchmark detector for all cases, with an SNR advantage up to 6 dB.



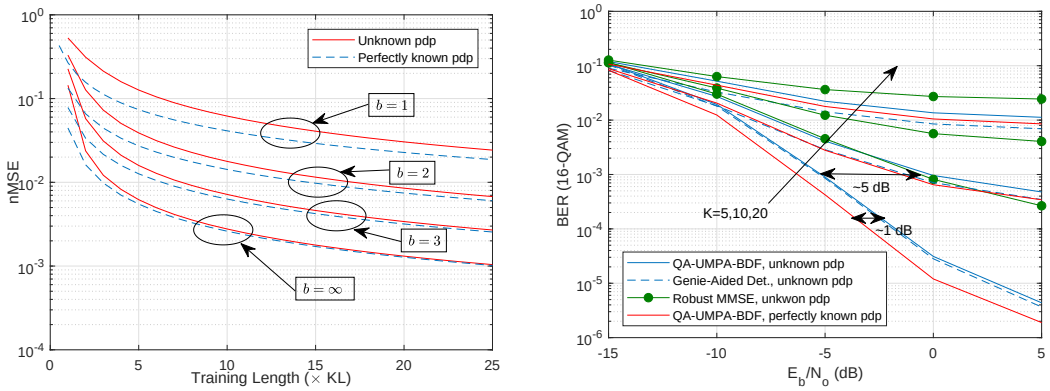
(a) BER vs. E_b/N_o, 16-QAM, $L = 32$, $q = 2$, correlated Rayleigh.

(b) BER vs. E_b/N_o, $L = 5$, 16-QAM, $q = 2$, correlated Rician.

Figure 7.4: BER vs. E_b/N_o for spatially correlated channel case.

Note that, we assumed perfect knowledge of the PDP of the channel in the channel estimation phase. The reason is that the channel coefficients $h'_{m,k}[\ell]$ change much faster due to small scale fading compared to the PDP of the channel, whose estimation is easier and do not require much overhead. In fact, it is shown that it is possible to estimate them without any additional pilots [23, 110]. That is the reason why they are widely assumed to be perfectly known in many different studies such as [23, 30]. However, we also want to demonstrate the effect of unknown PDP on the proposed channel estimation and detection performance by assuming a uniform PDP while constructing matrix $\mathbf{X}$ for the channel estimation using (7.25) although the actual PDP is COST-207 PDP. For this case, nMSE can be calculated analytically with an

expression similar to (7.53). The results in Fig. 7.2a (only for $K = 20$) and Fig. 7.3b (only for $K = 5, 10, 20$) are obtained for this unknown PDP case as in Fig. 7.5a, b. In Fig. 7.5a, we can see that training length must be increased a little, owing to



(a) nMSE vs. τ, $E_b/N_o = 0$ dB, $K = 20$, unknown PDP, Rayleigh.

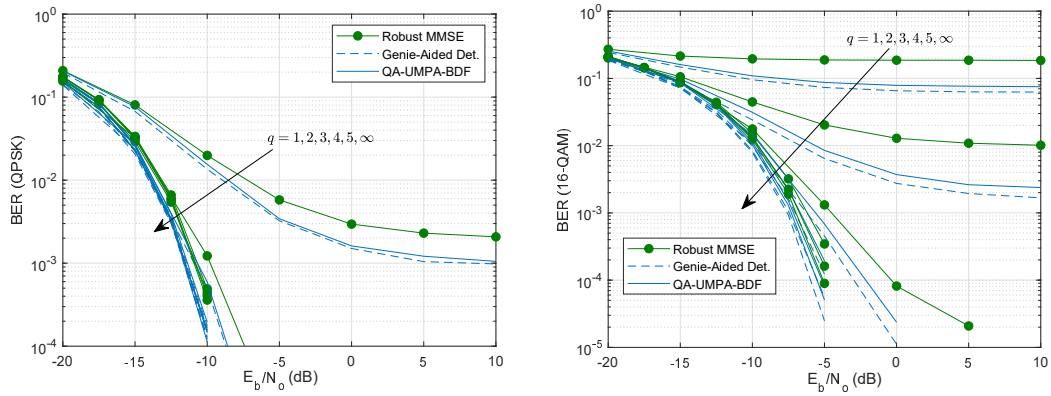
(b) BER vs. E_b/N_o for 16-QAM, $q = 2$, unknown PDP, Rayleigh.

Figure 7.5: Simulation results for unknown PDP.

the discrepancy between the assumed and the actual PDP, although the performance loss diminishes as the training length or the number of bits is increased. To see the performance loss in terms of E_b/N_o, we present the BER performances for 16-QAM and $q = 2$ in Fig. 7.5b. As noted in Fig. 7.5b, the performance loss due to unknown PDP is limited to ~ 1 dB, while the performance difference of ~ 5 dB between the benchmark detector and QA-UMPA-BDF observed in Fig. 7.3b is preserved.

As the next simulation scenario, we plot the error-rate performances for fixed K but varying q in Fig. 7.6. For all cases, QA-UMPA-BDF again has better performance. The performance gap between the two detectors is widened for 16-QAM. For QPSK and 16-QAM, performance improvement is not much for $q > 2$ and $q > 3$, respectively. Two iterations is again observed to be sufficient for QA-UMPA-BDF detector to match genie-aided detector performance.

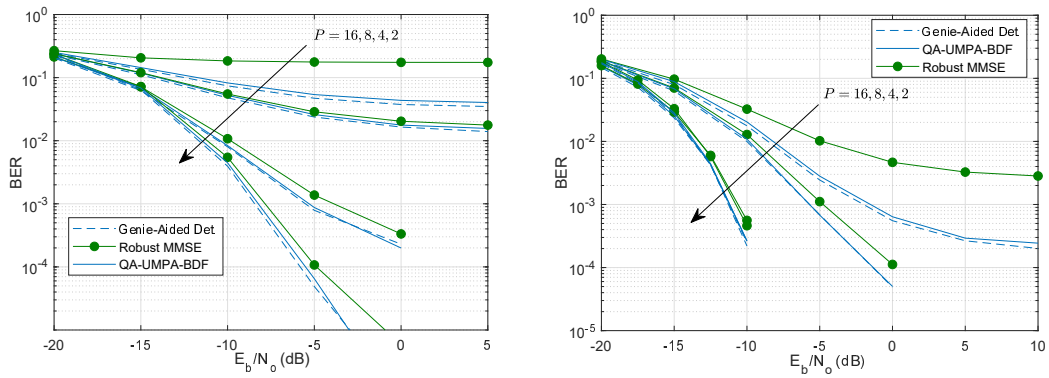
In the next simulation setting, the BER performances are observed for various modulation sizes (16-QAM, 8-PSK, 4-PSK, BPSK) in Fig.7.7. QA-UMPA-BDF again exhibits better performance compared to the benchmark detector for all modulation types, with significant performance difference for 16-QAM modulation. The reason to observe different BER for BPSK and QPSK is due to the correlation in the noise



(a) QPSK. (b) 16-QAM.

Figure 7.6: BER vs. SNR $K = 15$, $q = 1, 2, 3, 4, 5, \infty$ for QPSK (a), 16-QAM (b).

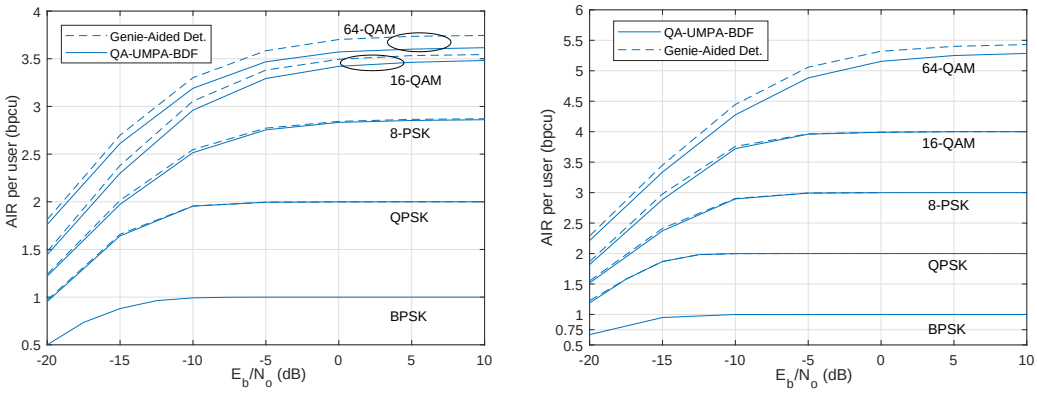
statistics stemming from the nonlinear quantizer. Such BER performance difference between BPSK and QPSK under quantization is also reported in [111].



(a) $q = 1$. (b) $q = 2$.

Figure 7.7: BER vs. SNR for $P = 16, 8, 4, 2$, $K = 10$, $q = 1$ (a) and $q = 2$ (b).

We also obtain per user AIR vs. SNR curves for $q = 1$ and $q = 2$ in Fig. 7.8. In Fig. 7.8a, QA-UMPA-BDF detector asymptotically provides an AIR about 2.8 bit per channel use (bpcu) for $q = 1$ with 8-PSK, close to the maximum AIR of 3 bpcu for 8-PSK. With 64-QAM, AIR can be asymptotically up to 3.5 bpcu for $q = 1$. For $q = 2$, we can see from Fig. 7.8b that up to 5.5 bpcu can be achieved with 64-QAM, close to the maximum AIR value of 6 bpcu for 64-QAM. This implies that a proper code with rate $5.5/6$ can provide very small BER values for 64-QAM.



(a) $q = 1$.

(b) $q = 2$.

Figure 7.8: Per user AIR vs. E_b/N_o, $K = 10$, $q = 1$ (a), $q = 2$ (b).

In the next simulation setting, we obtain the total AIR instead of per user AIR as a function of the number of users (K) in Fig. 7.9 when $E_b/N_o = 0$ dB, $M = 50$, $I = 7$, $L = 128$ with uniform PDP as a challenging ISI channel scenario. The key takeaways from Fig. 7.9 are as follows:

- Strong total AIR performance is observed with QPSK even with $q = 2$ and a very loaded case of 60 users, which is more than the number of antennas (maximum possible total AIR for 60 users is 120 bpcu with QPSK). Total AIR always rises with increasing K.
- For 8-PSK, total AIR is better than QPSK for all q, if not similar. For $q = 4$ maximum total AIR is achieved even with $K = 60$. Total AIR always rises with increasing K.
- For $q < 4$ with 16-QAM, total AIR always increase with K. For $q = 4, 5, \infty$, the maximum total AIR is observed for $K \approx 45, 53, 57$. For $q > 4$, total AIR increase with K if $K < 55$, which indicates a competent performance. Depending on the number of bits and users, 16-QAM has better total AIR performance than QPSK or 8-PSK in many cases.
- 64-QAM has superior total AIR performance compared to other modulation types for $q > 4$ and $K < 35$. For higher K and lower q, smaller modulation orders provide better total AIR in some cases. The maximum total AIR of about 200 bpcu is similar to that of 16-QAM.

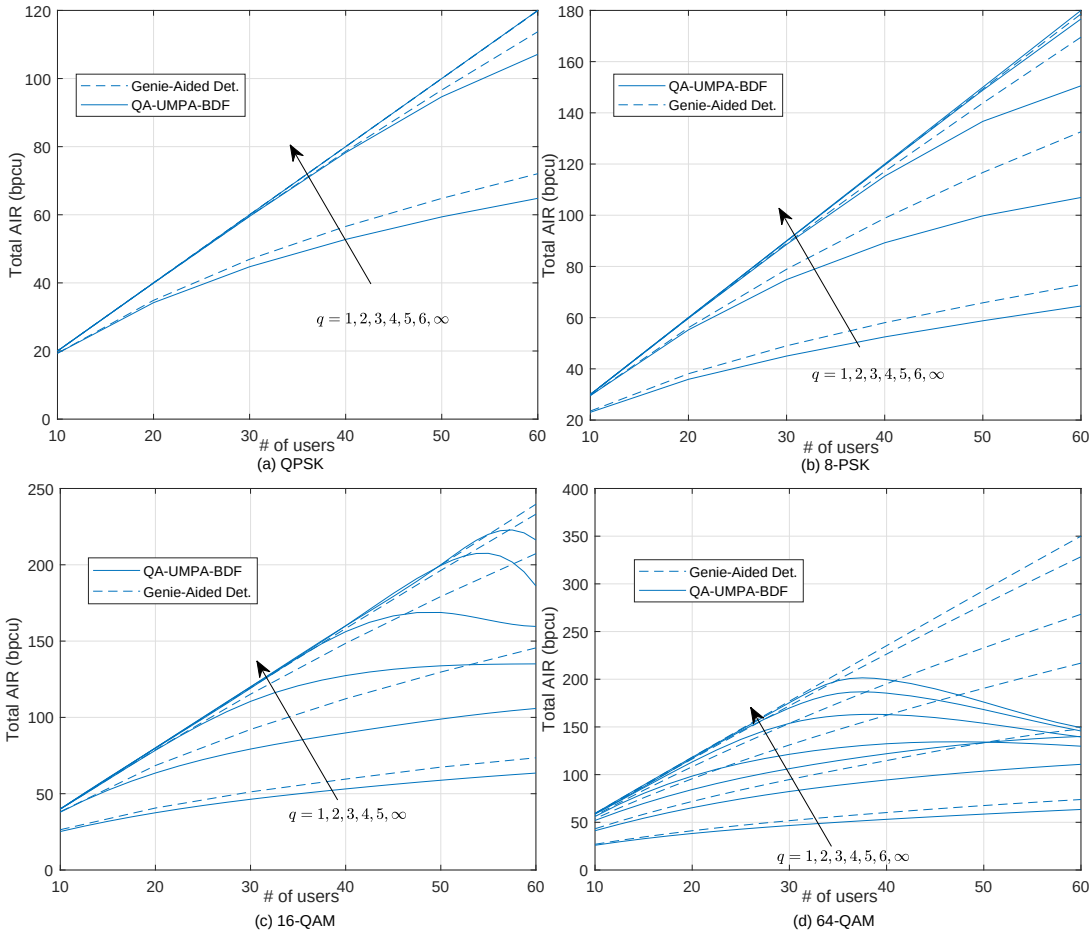
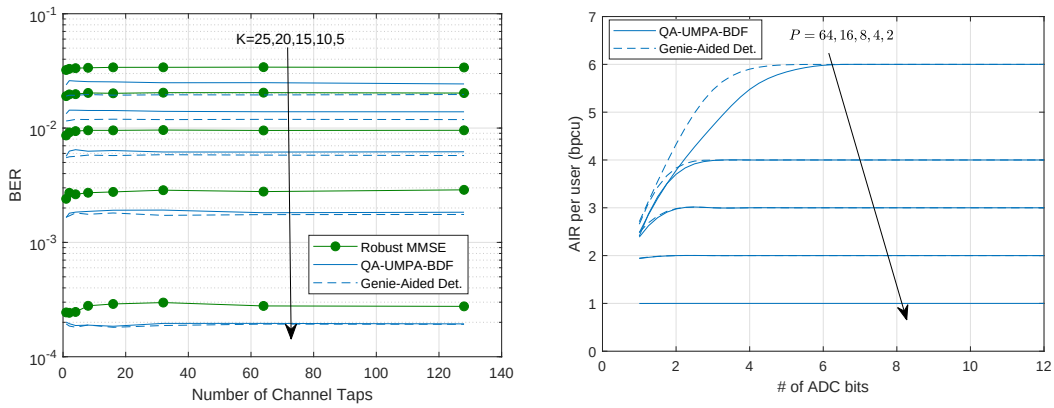


Figure 7.9: Total AIR vs. number of users, $M = 50$, $E_b/N_o = 0$ dB, $q = 1, 2, \ldots, 6, \infty$, QPSK (a), 8-PSK (b), 16-QAM (c), 64-QAM (d), $L = 128$, uniform PDP.

For the final simulation cases, we present the BER performance for $E_b/N_o = 0$ when the number of channel taps are varied in Fig. 7.10a. Moreover, AIR per user vs. number of ADC bits performance for $E_b/N_o = 0$ dB, $K = 25$, and $I = 7$ is presented in Fig. 7.10b. In Fig. 7.10a, it can be seen that the proposed QA-UMPA-BDF detector



(a) BER vs. L for $E_b/N_o = 0$ dB, $q = 1$, QPSK, uniform PDP.

(b) Per user AIR vs. number of bits q for $K = 25$, $E_b/N_o = 0$ dB.

Figure 7.10: BER vs. number of channel taps L (a) or per user AIR vs. number of bits q (b).

has a very robust BER performance to the changes in the number of channel taps. It can cancel ISI in time-domain effectively even when the number of channel taps is as large as 128, with no significant additional complexity (note that the complexity of QA-UMPA-BDF was increasing linearly with L when $J = 0$). It also always has better performance compared to the representative detector. From the results in Fig. 7.10b, it can be stated that 64-QAM can be employed with maximum possible AIR if $q > 5$, while the maximum possible AIR is achieved for $q > 2$ with 16-QAM. Moreover, as 64-QAM provides higher AIR per user values, it can be preferred to other modulation sizes for $K = 25$ when used with outer channel coding.

7.6 Conclusions

In this chapter, we proposed an LMMSE channel estimation and a low-complexity quantization-aware message passing detector based on bidirectional decision feedback. The proposed detector has very low complexity compared to the existing work

in the literature for highly dispersive channels with large number of channel taps, thanks to its reduced state sequence estimation capability. Under imperfect CSI, the proposed QA-UMPA-BDF detector is observed to outperform (significantly for some cases) a representative detector from the literature with comparable complexity but lower spectral efficiency due to its requirement to use CP. In short, we see that we are able to propose a detector than can perform better than the existing detectors in the literature having comparable complexity, even without resorting to any oversampling in time.

CHAPTER 8

CONCLUSION

In this thesis, the main objective is to find an answer to the question whether large number of antennas in massive MIMO is sufficient to obtain an adequate performance in terms of a reliable communication standpoint when low-resolution ADCs are employed. A related question is whether oversampling in time (or temporal oversampling) will provide a significant performance gain for massive MIMO systems with low-resolution ADCs.

To obtain answers for the above questions, we start with investigating the benefits of temporal oversampling in one-bit quantized massive MIMO systems with analytical tools and simulations. We show by deriving analytical performance bounds, whose accuracy is verified by simulations, that oversampling creates significant performance advantages when it is employed with one-bit quantized massive MIMO systems. Based on the results in this chapter, we state that having large number of antennas without resorting to any oversampling in time can be regarded as a missed opportunity since temporal oversampling enhance performance with low complexity.

In Chapter 4, we extend the work in Chapter 3 to frequency-selective channels as most practical channels of interest are frequency-selective. We see that the conclusions drawn for frequency-flat channels in Chapter 3 do not change for frequency-selective channels. We also observe that the advantages observed with the application of temporal oversampling are even more apparent for frequency-selective channels compared to frequency-flat channels.

Although the advantages observed when temporal oversampling is performed are remarkable, the ZF type detectors proposed in Chapter 3-4 are impractical, whose com-

plexity grows with the cube of block length. This may make the conclusions regarding the advantages of temporal oversampling drawn in Chapters 3-4 questionable. Therefore, we propose a low-complexity detector in Chapter 5 having linear complexity growth with the block length. Despite having much less complexity, the performance of the detector in Chapter 5 is very similar to the ZF type detectors in Chapter 3-4. Therefore, we illustrate in Chapter 5 that the advantages observed with temporal oversampling can also be obtained with a much lower complexity detector, making the conclusions associated with the advantages of temporal oversampling valid also for detectors with reasonable complexity.

Despite the fact that aforementioned advantages of temporal oversampling in time are promising for quantized massive MIMO systems, the question whether these advantages will be preserved when there is a source of significant interference from an adjacent band remains to be answered. In Chapter 6, we tried to find the answer to this question by making the performance analysis of quantized massive MIMO under adjacent channel interference when temporal oversampling is applied. Moreover, we employed OFDM modulation in that chapter, whereas single-carrier modulation was considered in the previous chapters. With the performance analysis presented in Chapter 6, whose accuracy is verified by simulations, we show that temporal oversampling provides significant gains in terms of compensating for any performance loss due to adjacent channel interference caused by low-resolution quantizers. Therefore, it is also meaningful to resort to the temporal oversampling technique also when significant adjacent channel interference is present in quantized massive MIMO.

In the aforementioned chapters, we try to find an answer whether oversampling in time has benefits for massive MIMO when significant quantization noise is present. We believe that we have obtained meaningful results supporting the fact that oversampling in time provides significant advantages. However, in Chapter 7, we question whether it is possible to obtain a detector that does not resort to any oversampling in time but provides superior performance compared to the existing detectors for quantized massive MIMO with comparable complexity. For that purpose, we propose a near optimal low-complexity factor-graph based detector. The proposed detector can work under frequency-selective channels and has a linear complexity growth with the number of channel taps, despite working with single-carrier modulation and not

requiring a cyclic-prefix. We show that the proposed detector can outperform the representative detectors in the literature in most of the investigated scenarios.

In summary, we investigated possible advantages that can be attained by temporal oversampling for quantized massive MIMO in this thesis. We have shown analytically and numerically that temporal oversampling can provide significant advantages in terms of error-rate and achievable rate performance of quantized massive MIMO without incurring significant additional computational complexity. Therefore, we state that temporal oversampling technique should always be considered in the design of massive MIMO systems with low-resolution quantizers. However, even without any oversampling, by making use of an efficient Ungerboeck type detector, we have shown that it is possible to obtain better performance compared to the existing detectors for quantized massive MIMO in the literature.

REFERENCES

[1] C. H. Sterling, *Military communications: from ancient times to the 21st century*. Santa Barbara, CA: ABC-CLIO, 2008.

[2] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge, UK: Cambridge Univ. Press, 2016.

[3] A. Goldsmith, *Wireless Communications*. New York, NY: Cambridge University Press, 2005.

[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] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *Proc. IEEE Int. Conf. Commun.*, London, 2015, pp. 1304–1309.

[6] L. Landau and G. Fettweis, "On reconstructable ASK-sequences for receivers employing 1-bit quantization and oversampling," in *Proc. IEEE Int. Conf. Ultra-Wideband*, Paris, 2014, pp. 180–184.

[7] B. Murmann, "The race for the extra decibel: A brief review of current ADC performance trajectories," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 58–66, Jul. 2015.

[8] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[9] ——, "Analog-to-digital converter technology comparison," in *IEEE GaAs IC Symp. Tech. Dig.*, Orlando, FL, 1994, pp. 217–219.

[10] B. Murmann. (2016, Jul.) *ADC Performance Survey 1997-2016*. [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[11] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May. 2013.

[12] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[13] I. D. O'Donnell and R. W. Brodersen, "An ultra-wideband transceiver architecture for low power, low rate, wireless systems," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1623–1631, Sept. 2005.

[14] S. Hoyos, B. M. Sadler, and G. R. Arce, "Monobit digital receivers for ultra-wideband communications," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1337–1344, Jul. 2005.

[15] S. Berger *et al.*, "Dynamic range-aware uplink transmit power control in LTE networks: Establishing an operational range for LTE's open-loop transmit power control parameters(α, p_0)," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 521–524, Oct. 2014.

[16] A. B. Üçüncü and A. Ö. Yılmaz, "Oversampling in one-bit quantized massive MIMO systems and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7952–7964, Dec. 2018.

[17] A. B. Üçüncü and A. Ö. Yılmaz, "Uplink performance analysis of oversampled wideband massive MIMO with one-bit ADCs," in *Proc. IEEE 88th Veh. Technol. Conf.*, Chicago, IL, Aug. 2018, pp. 1–5.

[18] A. B. Üçüncü and A. Ö. Yılmaz, "Sequential linear detection in one-bit quantized uplink massive MIMO with oversampling," in *Proc. IEEE 88th Veh. Technol. Conf.*, 2018, pp. 1–5.

[19] A. B. Üçüncü, E. Björnson, H. Johansson, A. Ö. Yılmaz, and E. G. Larsson, "Performance analysis of quantized uplink massive MIMO-OFDM with oversampling under adjacent channel interference," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 871–886, Feb. 2020.

[20] ——, "Performance of one-bit massive MIMO with oversampling under adjacent channel interference," in *Proc. IEEE Global Commun. Conf.*, 2019, pp. 1–6.

[21] A. B. Üçüncü, G. M. Güvensen, and A. Ö. Yılmaz, "A reduced complexity ungerboeck receiver for quantized wideband massive sc-mimo," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4921–4936, 2021.

[22] A. Ö. Yılmaz and A. B. Üçüncü, "Quantized detection in uplink MIMO with oversampling," U.S. Patent 10,447,504, Oct. 15, 2019.

[23] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[24] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, no. 3-4, pp. 154–655, 2017.

[25] S. Jacobsson, U. Gustavsson, G. Durisi, and C. Studer, "Massive MU-MIMO-OFDM uplink with hardware impairments: Modeling and analysis," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2018, pp. 1829–1835.

[26] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1595–1609, Mar. 2019.

[27] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Info. Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[28] J. Max, "Quantizing for minimum distortion," *IRE Trans. Info. Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[29] S. Jacobsson, "Massive multi-antenna communications with low-resolution data converters," Ph.D. dissertation, Dept. Elec. Eng., Chalmers Univ. Tech., Gothenburg, Sweden, 2019.

[30] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu,

"Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[31] Ö. T. Demir and E. Björnson. (2020, May 4) *The Bussgang Decomposition of Non-Linear Systems: Basic Theory and MIMO Extensions*. [Online]. Available: https://arxiv.org/pdf/2005.01597.pdf

[32] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Elec., Cambridge, MA, USA, Tech. Rep., Mar. 1952.

[33] H. E. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1519–1525, 1982.

[34] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[35] A. C. Ulusoy *et al.*, "A 60 ghz multi-gb/s system demonstrator utilizing analog synchronization and 1-bit data conversion," in *Proc. IEEE 13th Topical Meeting Silicon Monolithic Integrated Circuits RF Syst.*, Austin, TX, 2013, pp. 99–101.

[36] C. Risi, D. Persson, and E. G. Larsson. (2014, Apr. 30) *Massive MIMO with 1-bit ADC*. [Online]. Available: http://arxiv.org/abs/1404.7736

[37] B. M. Murray and I. B. Collings, "AGC and quantization effects in a zero-forcing MIMO wireless system," in *Proc. IEEE 63rd Veh. Technol. Conf.*, Melbourne, 2006, pp. 1802–1806.

[38] S. Bender *et al.*, "Communication with 1-bit quantization and oversampling at the receiver: Spectral constrained waveform optimization," in *Proc. 2016 IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun.*, 2016, pp. 1–5.

[39] J. A. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Proc. Int. Conf. Wireless Commun. Mobile Computing*, Vancouver, 2006, pp. 1387–1392.

[40] M. T. Ivrlac and J. A. Nossek, "Challenges in coding for quantized MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, 2006, pp. 2114–2118.

[41] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, 2007, pp. 1286–1289.

[42] J. Zhang *et al.*, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, May. 2016.

[43] Q. Bai and J. A. Nossek, "Energy efficiency maximization for 5G multi-antenna receivers," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 1, pp. 3–14, Oct. 2015.

[44] A. Mezghani *et al.*, "A modified MMSE receiver for quantized MIMO systems," *Proc. ITG/IEEE WSA*, 2007.

[45] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, 2012.

[46] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.

[47] M. Sarajlic *et al.*, "When are low resolution ADCs energy efficient in massive MIMO?" *IEEE Access*, vol. 5, pp. 14 837–14 853, Jul. 2017.

[48] A. Mezghani *et al.*, "An iterative receiver for quantized MIMO systems," in *Proc. 16th IEEE Mediterranean Electrotechnical Conf.*, Yasmine Hammamet, 2012, pp. 1049–1052.

[49] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.

[50] J. Choi *et al.*, "Quantized distributed reception for MIMO wireless systems using spatial multiplexing," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3537–3548, Jul. 2015.

[51] L. Landau *et al.*, "1-bit quantization and oversampling at the receiver: Communication over bandlimited channels with noise," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1007–1010, May. 2017.

[52] T. Halsig *et al.*, "Information rates for faster-than-nyquist signaling with 1-bit quantization and oversampling at the receiver," in *Proc. IEEE 79th Veh. Technol. Conf.*, 2014, pp. 1–5.

[53] J. Choi *et al.*, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May. 2016.

[54] H. Cramér, *Random variables and probability distributions*. New York, NY: Cambridge Univ. Press, 2004.

[55] D. Bertsekas and J. Tsitsiklis, *Introduction to probability*. Belmont, MA: Athena Scientific, 2008.

[56] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. New York, NY: McGraw-Hill Education, 2002.

[57] A. B. Üçüncü and A. Ö. Yılmaz, "Performance analysis of faster than symbol rate sampling in 1-bit massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.

[58] H. Q. Ngo, *Massive MIMO: Fundamentals and system designs*. Linköping, Sweden: Linköping Univ. Electron. Press, 2015.

[59] T. Zhang *et al.*, "Mixed-ADC massive MIMO detectors: Performance analysis and design optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Nov. 2016.

[60] *Universal Mobile Telecommunication System (UMTS); Base Station (BS) Radio Transmission and Reception (FDD)*, 3GPP TS 25.104, Release 12, 2015.

[61] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[62] *Evolved Universal Terrestrial Radio Access (E-UTRA) Physical channels and modulation*, 3GPP Std. TS 36.211 V8.9.0.

[63] H. E. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *BELL Syst. Tech. J.*, vol. 61, no. 7, pp. 1519–1525, 1982.

[64] J. H. V. Vleck and D. Middleton, "The spectrum of clipped noise," *Proc. IEEE*, vol. 54, no. 1, pp. 2–19, 1966.

[65] C. De Boor *et al.*, *A practical guide to splines*. New York: Springer-Verlag, 1978, vol. 27.

[66] I. . Lai *et al.*, "Asymptotic BER analysis for MIMO-BICM with zero-forcing detectors assuming imperfect CSI," in *2008 IEEE Int. Conf. Commun.*, May. 2008, pp. 1238–1242.

[67] S. Sumathi, P. Surekha, and P. Surekha, *LabVIEW based advanced instrumentation systems*. Berlin, Germany: Springer, 2007, vol. 728.

[68] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive MIMO with quantized phase-only measurements," in *Proc. Int. Conf. Commun.*, 2015, pp. 4576–4581.

[69] N. J. Myers and R. W. Heath, "Message passing-based joint CFO and channel estimation in mmWave systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3064–3077, Jun. 2019.

[70] H. G. Myung, J. Lim, and D. J. Goodman, "Peak-to-average power ratio of single carrier FDMA signals with pulse shaping," in *Proc. Int. Symp. Personal, Indoor Mobile Radio Commun.*, 2006, pp. 1–5.

[71] F. Pancaldi, G. M. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 37–56, Sep. 2008.

[72] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279, Aug. 2012.

[73] S. Buzzi, C. D'Andrea, T. Foggi, A. Ugolini, and G. Colavolpe, "Single-carrier modulation versus OFDM for millimeter-wave wireless MIMO," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1335–1348, Mar. 2018.

[74] X. Song, S. Haghighatshoar, and G. Caire, "Efficient beam alignment for millimeter wave single-carrier systems with hybrid MIMO transceivers," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1518–1533, Mar. 2019.

[75] J. Liu, Z. Luo, and X. Xiong, "Low-resolution ADCs for wireless communication: A comprehensive survey," *IEEE Access*, vol. 7, pp. 91 291–91 324, 2019.

[76] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jun. 2015.

[77] C. Cao, H. Li, and Z. Hu, "An AMP based decoder for massive MU-MIMO-OFDM with low-resolution ADCs," in *Proc. Int. Conf. Comput., Netw. Commun.*, 2017, pp. 449–453.

[78] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[79] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.

[80] Y. Wang, W. Xu, H. Zhang, and X. You, "Wideband mmWave channel estimation for hybrid massive MIMO with low-precision ADCs," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 285–288, Feb. 2019.

[81] L. V. Nguyen and A. L. Swindlehurst. (2020, Mar. 24) *SVM-based Channel Estimation and Data Detection for One-Bit Massive MIMO Systems*. [Online]. Available: https://arxiv.org/abs/2003.10678

[82] H. He, C. Wen, and S. Jin, "Bayesian optimal data detector for hybrid mmWave MIMO-OFDM systems with low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 469–483, Jun. 2018.

[83] S. Wang, L. Zhang, Y. Li, J. Wang, and E. Oki, "Multiuser MIMO communication under quantized phase-only measurements," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1083–1099, Mar. 2016.

[84] J. Garcia, J. Munir, R. Kilian, and J. A. Nossek. (2016, Sep. 15) *Channel estimation and data equalization in frequency-selective MIMO systems with one-bit quantization*. [Online]. Available: https://arxiv.org/abs/1609.04536

[85] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.

[86] J. Guerreiro, R. Dinis, and P. Montezuma, "Low-complexity SC-FDE techniques for massive MIMO schemes with low-resolution ADCs," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2368–2380, Mar. 2019.

[87] Y. Jeon, N. Lee, S. Hong, and R. W. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, Jul. 2018.

[88] Y. Jeon, N. Lee, and H. V. Poor, "Robust data detection for MIMO systems with one-bit ADCs: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1663–1676, Mar. 2020.

[89] J. Munir, D. Plabst, and J. A. Nossek, "Efficient equalization method for cyclic prefix-free coarsely quantized massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.

[90] M. Mohammadkarimi and M. Ardakani, "Optimal channel equalizer for mmWave massive MIMO using 1-bit ADCs in frequency-selective channels," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 882–885, Apr. 2020.

[91] Y. Jeon, H. Do, S. Hong, and N. Lee, "Soft-output detection methods for sparse millimeter-wave MIMO systems with low-precision ADCs," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2822–2836, Apr. 2019.

[92] J. Zhang, Z. Zheng, Y. Zhang, J. Xi, X. Zhao, and G. Gui, "3D MIMO for 5G NR: Several observations from 32 to massive 256 antennas based on channel measurement," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 62–70, 2018.

[93] O. Özdogan, E. Björnson, and E. G. Larsson, "Massive MIMO with spatially correlated rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, 2019.

[94] A. Kurt and G. M. Güvensen, "An adaptive hybrid beamforming scheme for time-varying wideband massive MIMO channels," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–7.

[95] T. Brown, E. D. Carvalho, and P. Kyritsi, *Practical guide to MIMO radio channel: With MATLAB examples*. Hoboken, NJ, USA: Wiley, 2012.

[96] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. Inf. Theory and Appl. Workshop*, San Diego, CA, 2015, pp. 191–198.

[97] Q. Bai and J. A. Nossek, "Energy efficiency maximization for 5G multi-antenna receivers," *Transactions Emerg. Telecommun. Technol.*, vol. 26, no. 1, pp. 3–14, Oct. 2014.

[98] A. Kipnis and G. Reeves, "Gaussian approximation of quantization error for estimation from compressed data," in *Proc. IEEE Int. Symp. Info. Theory*, 2019, pp. 2029–2033.

[99] J. E. Mazo, "Quantizing noise and data transmission," *Bell Syst. Tech. J.*, vol. 47, no. 8, pp. 1737–1753, Oct. 1968.

[100] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 446–472, Jul. 1948.

[101] W. Kester, "Taking the mystery out of the infamous formula, "SNR= 6.02n+ 1.76 dB" and why you should care," 2005, Analog Devices, MT-001.

[102] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, no. 5, pp. 624–636, May. 1974.

[103] G. Colavolpe, D. Fertonani, and A. Piemontese, "SISO detection over linear channels with linear complexity in the number of interferers," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1475–1485, Dec. 2011.

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

[105] M. Ivanov, C. Häger, F. Brännström, A. Graell i Amat, A. Alvarado, and E. Agrell, "On the information loss of the max-log approximation in BICM systems," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3011–3025, Jun. 2016.

[106] G. M. Güvensen, Y. Tanık, and A. Ö. Yılmaz, "A reduced-state ungerboeck type MAP receiver with bidirectional decision feedback for M-ary quasi orthogonal signaling," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 552–566, Feb. 2014.

[107] P. Viswanath *et al.*, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1968–1983, Sept. 1999.

[108] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sept. 1996.

[109] M. Salehi and J. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2007.

[110] T. Cui and C. Tellambura, "Power delay profile and noise variance estimation for OFDM," *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 25–27, 2006.

[111] U. H. Rizvi, G. J. M. Janssen, and J. H. Weber, "BER analysis of BPSK and QPSK constellations in the presence of ADC quantization noise," in *Proc. Asia-Pacific Conf. Commun.*, 2008, pp. 1–5.

[112] A. van den Bos, "Complex gradient and hessian," *Proc. IEE - Vision, Image Signal Process.*, vol. 141, no. 6, pp. 380–383, Dec. 1994.

[113] S. Guiasu and A. Shenitzer, "The principle of maximum entropy," *The mathematical intelligencer*, vol. 7, no. 1, pp. 42–48, Jan. 1985.

[114] S. Jacobsson *et al.*, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.

APPENDIX A

PROOFS IN CHAPTER 3

A.1 Proof of Lemma 1 and Lemma 2

In this part of the Appendix, $\sigma^2_{x'_m} \leq \mathrm{Var}(x'_m|x_m = 0, \mathbf{H})$ will be shown in Corollary 1 and the proof that if $\mathrm{Var}(x'_m|x_m = 0, \mathbf{H})$ is used in place of $\sigma^2_{x'_m}$ in (3.32) to find $f(x'_m|x_m, \mathbf{H})$ and $p(\hat{x}_m|x_m, \mathbf{H})$, and then (3.21) is utilized to find SER, this will constitute an upper bound on the exact SER value will be shown in Corollary 2.

Lemma 5: $\sigma^2_{x'_m}$ is a monotonic function of any element of the vectors $\mathrm{Re}(\boldsymbol{\mu}_\mathbf{y})$ and $\mathrm{Im}(\boldsymbol{\mu}_\mathbf{y})$, where $\boldsymbol{\mu_y} = \mathbf{E}[\mathbf{y}]$, when the any arbitrary element of the vectors $\mathrm{Re}(\boldsymbol{\mu}_\mathbf{y})$ and $\mathrm{Im}(\boldsymbol{\mu}_\mathbf{y})$ are in either $(0, \infty)$ or $(-\infty, 0)$.

Proof: If there is no such monotonicity as in the lemma statement, there will be two different values of $\boldsymbol{\mu_y}$, namely $\boldsymbol{\mu}^\mathbf{1}_\boldsymbol{y}$ and $\boldsymbol{\mu}^\mathbf{2}_\boldsymbol{y}$ such that $\sigma^2_{x'_m}(\boldsymbol{\mu}^\mathbf{1}_\boldsymbol{y}) = \sigma^2_{x'_m}(\boldsymbol{\mu}^\mathbf{2}_\boldsymbol{y})$ for the specified intervals for $\boldsymbol{\mu_y}$ in the lemma statement. From the mean value theorem this implies that the gradient of $\sigma^2_{x'_m}$ in (3.37) with respect to the mean vector $\boldsymbol{\mu_y}$ should be equal to zero for some $\mathbf{b_m}$. From (3.37), this requires

$$\mathbf{b}^T_m \nabla^1_{\boldsymbol{\mu_{y_p}}}(\boldsymbol{\Gamma}_\mathbf{rr})\mathbf{b}^*_m = 0 \tag{A.1}$$

and

$$\mathbf{b}^T_m \nabla^2_{\boldsymbol{\mu_{y_p}}}(\boldsymbol{\Gamma}_\mathbf{rr})\mathbf{b}^*_m = 0 \tag{A.2}$$

for all $p = 1,\ 2,\ \cdots,\ MN\beta$, where $\nabla^1_{\boldsymbol{\mu_{y_p}}}(\boldsymbol{\Gamma}_\mathbf{rr})$ and $\nabla^2_{\boldsymbol{\mu_{y_p}}}(\boldsymbol{\Gamma}_\mathbf{rr})$ are the matrices that are formed by taking the first and the second elements of the element-wise complex gradient vector of the matrix $\boldsymbol{\Gamma}_\mathbf{rr}$ with respect to vector $\boldsymbol{\mu_{y_p}} = [\mu^R_{y_p} + j\mu^I_{y_p}\ \ \mu^R_{y_p} - j\mu^I_{y_p}]$, where $\mu^R_{y_p} = \mathrm{Re}(\mathbf{E}[y_p])$ and $\mu^I_{y_p} = \mathrm{Im}(\mathbf{E}[y_p])$, respectively. In (A.1) and (A.2), there are a total of $2MN\beta$ equations to be satisfied since (A.1) and (A.2) should hold for

all $p = 1,\ 2,\ \cdots,\ MN\beta$. However, the solution space that consists of the possible values of the vector $\mathbf{b_m}$ has a dimension of at most L, L being the dimension of the sample space of $\mathbf{b_m}$ (note that the elements of $\mathbf{b_m}$ are functions of the channel matrix $\mathbf{H}$, thus they are random entities). In Lemma 6, it will be shown that there exists a p value such that all diagonal elements of the matrices $\nabla^1_{\boldsymbol{\mu_{y_p}}}(\mathbf{\Gamma_{rr}})$ or $\nabla^2_{\boldsymbol{\mu_{y_p}}}(\mathbf{\Gamma_{rr}})$ are nonzero for the specified interval for $\boldsymbol{\mu_y}$ in the lemma statement. In such a case, the dimension of the solution space that consists of the possible values of $\mathbf{b_m}$ will be less than L. Since the dimension of the sample space of $\mathbf{b_m}$ is L, which is greater than the dimension of the solution space, the set of possible values of $\mathbf{b_m}$ satisfying (A.1) and (A.2) has zero probability. Therefore, the lemma statement holds with probability one. ■

Lemma 6: There exists a p value such that all diagonal elements of the matrices $\nabla^1_{\boldsymbol{\mu_{y_p}}}(\mathbf{\Gamma_{rr}})$ and $\nabla^2_{\boldsymbol{\mu_{y_p}}}(\mathbf{\Gamma_{rr}})$ are nonzero for bounded and nonzero $\boldsymbol{\mu_y}$.

Proof: Consider the first diagonal element of the matrix $\mathbf{\Gamma_{rr}}$, namely $\sigma^2_{r_1}$, the variance of r_1, which is equal to $2 - E[r_1]E[r_1^*]$. Since $r_1 = r_1^R + jr_1^I$, the variance of r_1 becomes $2 - E[r_1^R + jr_1^I]E[r_1^R - jr_1^I] = 2 - E[(r_1^R)]^2 - E[(r_1^I)]^2$. Assuming that r_1^R and r_1^I have normal distribution, $E[(r_1^R)]$ can be found as

$$\begin{aligned}
E[(r_1^R)] &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2_{y_1^R}}} exp\left(-\frac{(y_1^R - \mu_{y_1^R})^2}{2\sigma^2_{y_1^R}}\right) dy_1^R \\
&- \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2_{y_1^R}}} exp\left(-\frac{(y_1^R - \mu_{y_1^R})^2}{2\sigma^2_{y_1^R}}\right) dy_1^R \\
&= Q\left(\frac{-\mu_{y_1^R}}{\sigma^2_{y_1^R}}\right) - \left(1 - Q\left(\frac{-\mu_{y_1^R}}{\sigma^2_{y_1^R}}\right)\right) \\
&= 2Q\left(\frac{-\mu_{y_1^R}}{\sigma^2_{y_1^R}}\right) - 1, \qquad \text{(A.3)}
\end{aligned}$$

where y_1^R is the real part of y_1, whose mean and variance are denoted as $\mu_{y_1^R}$ and $\sigma^2_{y_1^R}$. $E[(r_1^I)]$ can also be found similarly by replacing y_1^R terms by y_1^I in (A.3). Now consider the complex gradient $\nabla_{\boldsymbol{\mu_{y_p}}}(\sigma^2_{r_1}) = [\nabla^1_{\boldsymbol{\mu_{y_p}}}(\sigma^2_{r_1})\ \nabla^2_{\boldsymbol{\mu_{y_p}}}(\sigma^2_{r_1})]^T$ when $p = 1$. It can be found as [112]

$$\nabla_{\boldsymbol{\mu_{y_1}}}(\sigma^2_{r_1}) = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}\begin{bmatrix} \frac{\partial(\sigma^2_{r_1})}{\partial\mu^R_{y_1}} \\ \frac{\partial(\sigma^2_{r_1})}{\partial\mu^I_{y_1}} \end{bmatrix}. \tag{A.4}$$

Since $\sigma^2_{r_1} = 2 - \mathbf{E}[r^R_1]^2 - \mathbf{E}[r^I_1]^2$, (A.3) implies

$$\begin{aligned}\frac{\partial(\sigma^2_{r_1})}{\partial\mu^R_{y1}} &= \frac{-2\partial(E[(r^R_1)])}{\partial\mu^R_{y1}} + \frac{-2\partial(E[(r^I_1)])}{\partial\mu^R_{y1}} \\ &= -\frac{4}{\sqrt{2\pi\sigma^2_{y^R_1}}} exp\left(-\frac{\mu^2_{y^R_1}}{2\sigma^2_{y^R_1}}\right)\end{aligned} \tag{A.5}$$

since $\frac{-2\partial(E[(r^I_1)])}{\partial\mu_{y^R_1}} = 0$ and $\frac{\partial Q(x)}{\partial x} = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{x^2}{2}\right)$. Similarly, $\frac{\partial(\sigma^2_{r_1})}{\partial\mu^I_{y1}}$ can be obtained as

$$\frac{\partial(\sigma^2_{r_1})}{\partial\mu^I_{y1}} = -\frac{4}{\sqrt{2\pi\sigma^2_{y^I_1}}} exp\left(-\frac{\mu^2_{y^I_1}}{2\sigma^2_{y^I_1}}\right). \tag{A.6}$$

From (A.4), (A.5) and (A.6) it can be stated that $\nabla_{\boldsymbol{\mu_{y_1}}}(\sigma^2_{r_1}) \to 0$ if and only if both $\mu^R_{y_1}$ and $\mu^I_{y_1}$ goes to $\pm\infty$. The proof is essentially the same for the other diagonal elements of the matrix $\mathbf{\Gamma_{rr}}$. ■

Lemma 7: $\sigma^2_{x'_m}$ is a monotonically decreasing function of the real (imaginary) part of any arbitrary element of the bounded vector $\boldsymbol{\mu_y}$, namely $\boldsymbol{\mu^i_y}$, when $0 < \mathrm{Re}(\boldsymbol{\mu^i_y}) < \infty$ $(0 < \mathrm{Im}(\boldsymbol{\mu^i_y}) < \infty)$ or a monotonically increasing function of the real (imaginary) part of $\boldsymbol{\mu^i_y}$ when $-\infty < \mathrm{Re}(\boldsymbol{\mu^i_y}) < 0$ $(-\infty < \mathrm{Im}(\boldsymbol{\mu^i_y}) < 0)$

Proof: Let $\mathrm{Re}(\boldsymbol{\mu^i_y}) \to \pm\infty$. This requires $\mathbf{E}[|x_m|] \to \infty$ since $\mathbf{E}[x_k] = 0$ for $k \neq m$, and the noise is also zero mean. In such case, the observation vectors $\mathbf{y}$ and $\mathbf{r}$ become deterministic since the interference due to x_k $k \neq m$, which corresponds to ISI and MUI, and the receiver noise becomes insignificant. Therefore, $\sigma^2_{x'_m} \to 0$. Similarly, $\sigma^2_{x'_m} \to 0$ also when $\mathrm{Im}(\boldsymbol{\mu^i_y}) \to \pm\infty$. The proof holds for all i values. This implies that $\sigma^2_{x'_m}(\boldsymbol{\mu_y} = \mathbf{0}) \geq \sigma^2_{x'_m}(\mathrm{Re}(\boldsymbol{\mu^i_y}) \to \pm\tilde{\infty})$ or $\sigma^2_{x'_m}(\boldsymbol{\mu_y} = \mathbf{0}) \geq \sigma^2_{x'_m}(\mathrm{Im}(\boldsymbol{\mu^i_y}) \to \pm\tilde{\infty})$ $\forall i$. Along with Lemma 5, this proves the lemma statement. ■

Corollary 1: $\sigma^2_{x'_m} \leq \mathrm{Var}(x'_m | x_m = 0, \mathbf{H})$

Proof: Due to Lemma 7, the maximum value of $\sigma^2_{x'_m}$ is at the point where $\boldsymbol{\mu_y} = \mathbf{0}$ which is only possible for given $x_m = 0$ case. ■

Corollary 2: If $\mathrm{Var}(x'_m|x_m = 0, \mathbf{H})$ *is used in place of* $\sigma^2_{x'_m}$ *the SER value calculated using (3.21) will yield an upper bound on the actual SER.*

Proof: Since $\mathrm{Var}(x'_m|x_m = 0, \mathbf{H}) \geq \sigma^2_{x'_m}$ due to Corollary 1, it follows that $\mathrm{Var}(\hat{x}_m|x_m = 0, \mathbf{H}) > \mathrm{Var}(\hat{x}_m|x_m, \mathbf{H})$. Therefore, when $\mathrm{Var}(\hat{x}_m|x_m = 0, \mathbf{H})$ is used in place of $\mathrm{Var}(\hat{x}_m|x_m, \mathbf{H})$, the values calculated for $p(\hat{x}_m|x_m, \mathbf{H})$ in (3.21) when $\hat{x}_m \neq x_m$ will be higher than their actual values, which will result in a higher SER than the actual SER. ■

A.2 The Details to Obtain $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$

To begin with, $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$ can be expressed as

$$\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = \begin{bmatrix} \mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} & \mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} \\ \mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} & \mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} \end{bmatrix}, \tag{A.7}$$

where $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} = \mathbf{E}[\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})^H]$, $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} = \mathbf{E}[\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})^H]$, $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} = \mathbf{E}[\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})^H]$ and $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} = \mathbf{E}[\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})^H]$. $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})}$ in (A.7) can be obtained as

$$\begin{aligned} \mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} =& \mathbf{E}\left[\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})^H\right] \\ =& \mathbf{E}\left[\frac{1}{2}\left(\mathbf{H}^*\mathbf{x}^* + \mathbf{n}^* + \mathbf{H}\mathbf{x} + \mathbf{n}\right)\right. \\ & \left.\times \frac{1}{2}\left(\mathbf{H}^*\mathbf{x}^* + \mathbf{n}^* + \mathbf{H}\mathbf{x} + \mathbf{n}\right)^H\right] \\ =& \frac{1}{4}\left[\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{H}^*\mathbf{R}^*_{\mathbf{xx}}\mathbf{H}^T + \mathbf{R}_{\mathbf{nn}} + \mathbf{R}_{\mathbf{nn}}{}^*\right], \end{aligned} \tag{A.8}$$

where $\mathbf{R}_{\mathbf{xx}} = \mathrm{diag}(1, \cdots, 1, 0, 1, \cdots, 1)$ is an $NK \times NK$ matrix whose m^{th} diagonal element is taken to be zero since x_m is assumed to be zero to find the upper bound on $\sigma^2_{x'_m}$. Moreover, $\mathbf{R}_{\mathbf{nn}}$ is the noise correlation matrix which can be expressed as

$$\mathbf{R}_{\mathbf{nn}} = \begin{bmatrix} \mathrm{diag}(\sigma_n^2, \cdots, \sigma_n^2) & \mathbf{R}_{Nos} \\ \mathbf{R}^*_{Nos} & \mathrm{diag}(\sigma_n^2, \cdots, \sigma_n^2) \end{bmatrix}_{MN\beta \times MN\beta}, \tag{A.9}$$

where

$$\mathbf{R}_{Nos} = \sigma_n^2 \begin{bmatrix} diag((\alpha_0^1)^2)_M & diag((\alpha_{-1}^1)^2)_M & \cdots & diag((\alpha_{-(N-1)}^1)^2)_M \\ diag((\alpha_1^1)^2)_M & diag((\alpha_0^1)^2)_M & \cdots & diag((\alpha_{-(N-2)}^1)^2)_M \\ diag((\alpha_2^1)^2)_M & diag((\alpha_1^1)^2)_M & \cdots & diag((\alpha_{-(N-3)}^1)^2)_M \\ \vdots & \ddots & \ddots & \vdots \\ diag((\alpha_{(N-1)}^1)^2)_M & diag((\alpha_{N-2}^1)^2)_M & \cdots & diag((\alpha_0^1)^2)_M \end{bmatrix}, \quad \text{(A.10)}$$

for $\beta = 2$, which can be found similarly for arbitrary β. In (A.10), $\alpha_i^m = g(iT + mT/\beta)$ and $diag((\alpha_i^1)^2)_M$ is an $M \times M$ diagonal matrix whose diagonal elements are equal to the square of α_i^1. Similarly, $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})}$, $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})}$ and $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})}$ can be found as

$$\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} = \frac{1}{4}\left[\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{H}^*\mathbf{R}_{\mathbf{xx}}^*\mathbf{H}^T + \mathbf{R}_{\mathbf{nn}} + \mathbf{R}_{\mathbf{nn}}^*\right], \quad \text{(A.11)}$$

$$\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} = \frac{-1}{4j}\left[\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H - \mathbf{H}^*\mathbf{R}_{\mathbf{xx}}^*\mathbf{H}^T + \mathbf{R}_{\mathbf{nn}} - \mathbf{R}_{\mathbf{nn}}^*\right], \quad \text{(A.12)}$$

$$\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} = \frac{1}{4j}\left[\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H - \mathbf{H}^*\mathbf{R}_{\mathbf{xx}}^*\mathbf{H}^T + \mathbf{R}_{\mathbf{nn}} - \mathbf{R}_{\mathbf{nn}}^*\right]. \quad \text{(A.13)}$$

Using (A.8)-(A.13), the matrix on the right-hand side (RHS) of (A.7) can be found to obtain $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}$. After that, (3.39) can be employed to find $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}$. To be able to employ (3.37), $\mathbf{\Gamma}_{\mathbf{rr}}$ needs to be found which is equal to $\mathbf{R}_{\mathbf{rr}}$ for the zero mean case, for which we find an upper bound on $\sigma_{x_m'}^2$. In order to obtain $\mathbf{R}_{\mathbf{rr}}$ from $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}$ we express $\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}}$ as

$$\mathbf{R}_{\tilde{\mathbf{r}}\tilde{\mathbf{r}}} = \begin{bmatrix} \mathbf{R}_{\mathrm{Re}(\mathbf{r})\,\mathrm{Re}(\mathbf{r})} & \mathbf{R}_{\mathrm{Re}(\mathbf{r})\,\mathrm{Im}(\mathbf{r})} \\ \mathbf{R}_{\mathrm{Im}(\mathbf{r})\,\mathrm{Re}(\mathbf{r})} & \mathbf{R}_{\mathrm{Im}(\mathbf{r})\,\mathrm{Im}(\mathbf{r})} \end{bmatrix}. \quad \text{(A.14)}$$

Since $\mathbf{r} = \mathrm{Re}(\mathbf{r}) + j\,\mathrm{Im}(\mathbf{r})$, $\mathbf{R}_{\mathbf{rr}}$ can be found as

$$\begin{aligned} \mathbf{R}_{\mathbf{rr}} =& \mathbf{R}_{\mathrm{Re}(\mathbf{r})\,\mathrm{Re}(\mathbf{r})} + \mathbf{R}_{\mathrm{Im}(\mathbf{r})\,\mathrm{Im}(\mathbf{r})} \\ & - j\mathbf{R}_{\mathrm{Re}(\mathbf{r})\,\mathrm{Im}(\mathbf{r})} + j\mathbf{R}_{\mathrm{Im}(\mathbf{r})\,\mathrm{Re}(\mathbf{r})}. \end{aligned} \quad \text{(A.15)}$$

The terms on the right-hand side of $(A.15)$ can be found from (A.14).

A.3 Proof of Lemma 3

Denote $p(\hat{x}_m = x_m | x_m, \mathbf{H}) = p_m$. When $H(x|y)$ denotes the conditional entropy of x conditioned on y, by definition of mutual information

$$I(x_m; \hat{x}_m) = \mathbf{E_H}\left[H(\hat{x}_m|\mathbf{H}) - H(\hat{x}_m|x_m, \mathbf{H})\right]. \tag{A.16}$$

Given that x_m is transmitted, there are three cases that $\hat{x}_m \neq x_m$ owing to QPSK type modulation. Denote the probabilities of these three events as p_{me1}, p_{me2} and p_{me3}. Therefore, $p_m + p_{me1} + p_{me2} + p_{me3} = 1$. Consider $H(\hat{x}_m|x_m, \mathbf{H})$. When we use the method of Lagrange multipliers to maximize $H(\hat{x}_m|x_m, \mathbf{H})$ with respect to p_m, p_{me1}, p_{me2} and p_{me3}, the only solution that makes the gradient of the Lagrangian zero is the case when $p_m = p_{me1} = p_{me2} = p_{me3} = 1/4$, which is the global maximum point of $H(\hat{x}_m|x_m, \mathbf{H})$ [113]. For a given x_m case that is nonzero and finite, this can only occur when $\sigma^2_{x'_m} \to \infty$. Since there is no other point that can make the gradient of the Lagrangian with respect to the vector $[p_m\ p_{me1}\ p_{me2}\ p_{me3}]$ zero, there is a unique local maximum of $H(\hat{x}_m|x_m, \mathbf{H})$ which occurs only when $\sigma^2_{x'_m} \to \infty$. Therefore, $H(\hat{x}_m|x_m, \mathbf{H})$ must be an increasing function of $\sigma^2_{x'_m}$. Moreover, $H(\hat{x}_m|\mathbf{H})$ in (A.16) does not depend on $\sigma^2_{x'_m}$ because $\hat{x}_m$ conditioned on $\mathbf{H}$ will be zero mean since there is no conditioning on x_m, thus it can take any QPSK symbol value with equal probability regardless of the value of $\sigma^2_{x'_m}$. Hence, it follows that $I(x_m; \hat{x}_m)$ is a decreasing function of $\sigma^2_{x'_m}$. Therefore, if we replace $\sigma^2_{x'_m}$ by $\mathrm{Var}(x'_m|x_m = 0, \mathbf{H})$, which is greater than the actual $\sigma^2_{x'_m}$, $I(x_m; \hat{x}_m)$ that we find will be lower than the actual $I(x_m; \hat{x}_m)$.

A.4 Proof of Lemma 4

$SNR \to 0$ requires $\sigma_n^2 \to \infty$ when the average transmitted power of the users is nonzero. In this case, according to (A.8) and (A.11-A.13), $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} \approx \mathbf{R_{nn}} + \mathbf{R_{nn}}^*$, $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} \approx \mathbf{R_{nn}} + \mathbf{R_{nn}}^*$, $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})} \approx \mathbf{R_{nn}} - \mathbf{R_{nn}}^*$ and $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})} \approx \mathbf{R_{nn}} - \mathbf{R_{nn}}^*$, none of which depends on the given value of x_m. Therefore, since $\mathbf{R_{\tilde{y}\tilde{y}}}$ depends only on $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})}$, $\mathbf{R}_{\mathrm{Re}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})}$, $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Re}(\mathbf{y})}$ and $\mathbf{R}_{\mathrm{Im}(\mathbf{y})\,\mathrm{Im}(\mathbf{y})}$, it also does not depend on the given value of x_m when $SNR \to 0$. Hence it follows that $\mathbf{R_{\tilde{r}\tilde{r}}}$, thus $\mathbf{R_{rr}}$ does not depend on x_m from (3.39). Moreover, when $\sigma_n^2 \to \infty$, this will result

in $\kappa_{1p} \to 0.25$, $\kappa_{2p} \to 0.25$, $\kappa_{3p} \to 0.25$, $\kappa_{4p} \to 0.25$, $\forall p$ according to (3.28). In this case, $\mu_p \to 0$ owing to (3.27) $\forall p$. This implies that $\mathbf{E}[\mathbf{y}] \to 0$, thus $\mathbf{E}[\mathbf{r}] \to 0$. When $\mathbf{E}[\mathbf{r}] \to 0$, $\mathbf{\Gamma_{rr}} \to \mathbf{R_{rr}}$. Therefore, since $\mathbf{R_{rr}}$ does not depend on the given value of x_m, $\mathbf{\Gamma_{rr}}$ also does not, which results in $\sigma^2_{x'_m}$ being independent of the given value of x_m according to (3.37) when $SNR \to 0$.

APPENDIX B

PROOFS IN CHAPTER 6

B.1 Proof of Proposition 2

Direct computation yields

$$\mathbf{C}_{\mathbf{r}}[m] = \mathbf{E}\left[\mathbf{r}[n]\mathbf{r}[n-m]^H\right] = \mathbf{E}\left[\frac{1}{N^2}\sum_{u=0}^{N-1}\sum_{u'=0}^{N-1}\tilde{\mathbf{r}}[u]\tilde{\mathbf{r}}[u']^H e^{j2\pi nu/N} e^{-j2\pi(n-m)u'/N}\right] \tag{B.1}$$

$$= \frac{1}{N^2}\sum_{u=0}^{N-1}\mathbf{E}\left[\tilde{\mathbf{r}}[u]\tilde{\mathbf{r}}[u]^H\right] e^{j2\pi nu/N} e^{-j2\pi(n-m)u/N} \tag{B.2}$$

$$= \frac{1}{N^2}\sum_{u=0}^{N-1}\mathbf{C}_{\tilde{\mathbf{r}}[u]} e^{j2\pi mu/N}. \tag{B.3}$$

Here (B.1) holds by definition, (B.2) is because $\mathbf{E}\left[\tilde{\mathbf{r}}[u]\tilde{\mathbf{r}}[u']^H\right] = \mathbf{0}$ for $u \neq u'$ as the data symbols of the users are independent and (B.3) also holds by definition. ■

B.2 Proof of Proposition 3

$\mathbf{C}_{\mathbf{q}[u]}$ can be found from $\mathbf{C}_{\mathbf{q}}[m]$ as

$$\mathbf{C}_{\mathbf{q}[u]} = \mathbf{E}\left[\mathbf{q}[u]\mathbf{q}[u]^H\right] = \sum_{m=0}^{N-1}\sum_{m'=0}^{N-1} \mathbf{E}\left[\tilde{\mathbf{q}}[m]\tilde{\mathbf{q}}[m']^H\right] e^{-j2\pi mu/N} e^{j2\pi m'u/N} \tag{B.4}$$

$$= \sum_{m=0}^{N-1}\sum_{m'=0}^{N-1} \mathbf{C}_{\mathbf{q}}[m-m'] e^{-j2\pi(m-m')u/N} \tag{B.5}$$

$$= \sum_{\ell=-(N-1)}^{N-1} (N-|\ell|)\mathbf{C}_{\mathbf{q}}[\ell] e^{-j2\pi\ell u/N} \tag{B.6}$$

$$= \left[\sum_{\ell=0}^{N-1}(N-\ell)\mathbf{C}_{\mathbf{q}}[\ell]e^{-j2\pi\ell u/N} + \sum_{\ell=0}^{N-1}(N-\ell)\mathbf{C}_{\mathbf{q}}[\ell]^* e^{j2\pi\ell u/N}\right] \tag{B.7}$$

$$= \mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[\ell]\right\} + \mathrm{DFT}_{N,u}\left\{\mathbf{\Gamma}[\ell]\right\}^* - N\mathbf{C}_{\tilde{\mathbf{q}}}[0]. \tag{B.8}$$

Here (B.4),(B.5) hold by definition and due to stationarity of $\mathbf{q}[m]$. A change of variable ($\ell = m - m'$) is introduced in (B.6). (B.7) holds since $\mathbf{C}_{\mathbf{q}}[\ell]^* = \mathbf{C}_{\mathbf{q}}[-\ell]$. ■

B.3 Proof of Proposition 4

When $\gamma_k[u]$ is as defined in (6.26), the fact that R in (6.25) is a lower bound ergodic capacity follows from [2, eq. (2.46)], since the conditions $\mathbf{E}[w_k[u]|\hat{\mathbf{H}}] = \mathbf{E}[\tilde{s}_k[u]^* w_k[u]|\hat{\mathbf{H}}] = \mathbf{E}[\hat{g}_k[u]'\tilde{s}_k[u]^* w_k[u]|\hat{\mathbf{H}}] = 0$, where $w_k[u] \triangleq i_k[u]+n_k[u]+q_k[u]$, hold for LMMSE channel estimate $\hat{\mathbf{H}}$. $\mathbf{E}[g_k[u]'|\hat{\mathbf{H}}]$ in the numerator of $\gamma_k[u]$ expression can be found as follows:

$$\mathbf{E}\left[g_k[u]'|\hat{\mathbf{H}}\right] = \mathbf{E}\left[\rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{h}}_k[u]||\hat{\mathbf{H}}\right] \tag{B.9}$$

$$= \mathbf{E}\left[\rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_k[u]||\hat{\mathbf{H}}\right] + \mathbf{E}\left[\rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{e}}_k[u]||\hat{\mathbf{H}}\right] \tag{B.10}$$

$$= \rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_k[u], \tag{B.11}$$

where the last step is owing to the fact that LMMSE channel estimates are unbiased thus the estimation errors are zero-mean. Var $\left[\hat{g}_k[u]'|\hat{\mathbf{H}}\right]$ can also be found as follows:

$$\text{Var}\left[\hat{g}_k[u]'|\hat{\mathbf{H}}\right] = \text{Var}\left[\rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_k[u] + \rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{e}}_k[u]|\hat{\mathbf{H}}\right] \tag{B.12}$$

$$= \text{Var}\left[\rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{e}}_k[u]|\hat{\mathbf{H}}\right] \tag{B.13}$$

$$= \rho_d^2 N\hat{\mathbf{b}}_k[u]^H\mathbf{A}\mathbf{E}\left[\tilde{\mathbf{e}}_k[u]\tilde{\mathbf{e}}_k[u]^H\right]\mathbf{A}^H\hat{\mathbf{b}}_k[u] \tag{B.14}$$

$$= \rho_d^2 N\sigma_e^2||\hat{\mathbf{b}}_k[u]^H\mathbf{A}||^2, \tag{B.15}$$

where (B.15) follows from the fact that the LMMSE channel estimation errors are uncorrelated with the channel estimates, hence with the ZF matrix row vectors. Moreover,

$$\text{Var}[i_k[u] + n_k[u] + q_k[u]|\hat{\mathbf{H}}] = \text{Var}\left[i_k[u]|\hat{\mathbf{H}}\right] + \text{Var}\left[n_k[u]|\hat{\mathbf{H}}\right] + \text{Var}\left[q_k[u]|\hat{\mathbf{H}}\right]. \tag{B.16}$$

(B.16) is due to uncorrelatedness of the thermal and quantization noise with every other term. The uncorrelatedness of the quantization noise with the other terms is due to the Bussgang decomposition; matrix $\mathbf{A}$ is selected to make $\mathbf{r}[n]$ to be uncorrelated with $\mathbf{q}[n]$ in (6.12) and this also holds for the frequency domain terms as DFT is a unitary transformation which preserves inner products. Moreover, Var $\left[i_k[u]|\hat{\mathbf{H}}\right]$ can be derived as follows:

$$\text{Var}\left[i_k[u]|\hat{\mathbf{H}}\right] \tag{B.17}$$

$$= \rho_d^2 N\text{Var}\left[\sum_{z\neq k, z\in\mathcal{K}_d}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_z[u]\tilde{s}_z[u] + \sum_{z\neq k, z\in\mathcal{K}_d}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{e}}_z[u]\tilde{s}_z[u]\middle|\hat{\mathbf{H}}\right] \tag{B.18}$$

$$= \rho_d^2 N\left[\sum_{z\neq k, z\in\mathcal{K}_d}|\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_z[u]|^2 + \sum_{z\neq k, z\in\mathcal{K}_d}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\mathbf{E}\left[\tilde{\mathbf{e}}_z[u]\tilde{\mathbf{e}}_z[u]^H\right]\mathbf{A}^H\hat{\mathbf{b}}_k[u]\right] \tag{B.19}$$

$$= \rho_d^2 N\left[\sum_{z\neq k, z\in\mathcal{K}_d}|\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_z[u]|^2 + \sum_{z\neq k, z\in\mathcal{K}_d}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\sigma_e^2\mathbf{A}^H\hat{\mathbf{b}}_k[u]\right] \tag{B.20}$$

$$= \rho_d^2 N\left[\sum_{z\neq k, z\in\mathcal{K}_d}|\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_z[u]|^2 + \sigma_e^2(K-1)||\hat{\mathbf{b}}_k[u]^H\mathbf{A}||^2\right] \tag{B.21}$$

where (B.19) is follows from the fact that the channel estimation errors for all users are also uncorrelated with each other and with ZF matrix row vectors. Moreover,

(B.20) is due owing to the fact that estimation errors for all channel coefficients are the same due to (6.43). It is also straightforward to show that Var $\left[n_k[u]|\hat{\mathbf{H}}, \tilde{s}_k[u]\right] = N_k[u]$, Var $\left[q_k[u]|\hat{\mathbf{H}}, \tilde{s}_k[u]\right] = Q_k[u]$. ■

B.4 Proof of Proposition 5

Denote the mth element of $\mathbf{C_r}[0]$ by $\mathbf{C_r}[0]_{m,m}$, which can be calculated as

$$\mathbf{C_r}[0]_{m,m} = \mathbf{E}\left[\left|\sum_{\ell=0}^{L-1}\sum_{k=1}^{K+I} h_{m,k}[\ell]s_k[n-l] + w_m[n]\right|^2 \middle| \underline{\mathbf{H}}\right] \tag{B.22}$$

$$= \sum_{\ell=0}^{L-1}\sum_{k=1}^{K+I} G_k|h_{m,k}[\ell]|^2 + N_o, \tag{B.23}$$

$G_k = (|\mathcal{U}_D|\rho_d^2)/N$ if $k \in \mathcal{K}_D$ or $G_k = (|\mathcal{U}_I|\rho_i^2)/N$ if $k \in \mathcal{K}_I$. Owing to Chebyshev inequality,

$$\Pr\left[|\mathbf{C_r}[0]_{m,m} - \mu| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\mathbf{C_r}[0]_{m,m}]}{\epsilon^2}\ \forall m, \tag{B.24}$$

where $\mu \triangleq \mathbf{E}_{\underline{\mathbf{H}}}\left[\mathbf{C_r}[0]_{m,m}\right] = (|\mathcal{U}_D|K\rho_d^2 + |\mathcal{U}_I|I\rho_i^2)/N + N_o$, $\Pr[\xi]$ denote the probability of an event ξ and $\mathrm{Var}[\mathbf{C_r}[0]_{m,m}]$ is the variance of $\mathbf{C_r}[0]_{m,m}$ with respect to $\underline{\mathbf{H}}$, which is equal to $3G_k^2/L$. Since $\mathrm{Var}[\mathbf{C_r}[0]_{m,m}]$ is decreasing with L, for any $\epsilon > 0$ we can find $L > 0$ such that $\Pr\left[|\mathbf{C_r}[0]_{m,m} - \mu| \geq \epsilon\right] = 0$, thus $\mathbf{C_r}[0]_{m,m}$ converge in probability to μ. Therefore, the error in the approximation $\mathbf{C_r}[0]_{m,m} \approx (|\mathcal{U}_D|K\rho_d^2 + |\mathcal{U}_I|I\rho_i^2)/N + N_o$ converge to zero in probability as L grows large (goes to infinity). This implies that

$$\mathbf{A} = \frac{2}{\sqrt{\pi}}\mathrm{diag}\left(\mathbf{C_r}[0]\right)^{-0.5}$$
$$\rightarrow \frac{2}{\sqrt{\pi}}\left(\left(|\mathcal{U}_D|K\rho_d^2 + |\mathcal{U}_I|I\rho_i^2\right)/N + N_o\right)^{-0.5}\mathbf{I} = G\mathbf{I}, \tag{B.25}$$

as L grows large (convergence is represented by $\rightarrow$ symbol). It can similarly be shown that $\mathbf{C_r}[0]_{m,i} \approx 0$ for $m \neq i$ as L grows large since $\mathbf{E}_{\underline{\mathbf{H}}}\left[\mathbf{C_r}[0]_{m,i}\right] = 0$ as $h_{m,k}[l]$, $h_{i,k}[l]$, $w_m[n]$ and $w_i[n]$ are uncorrelated for $m \neq i$. The numerator of the

$\gamma_k[u]'$ expression can be found as

$$|\mathbf{E}\left[g_k[u]'\right]|^2 = \left|\mathbf{E}\left[\rho_d\sqrt{N}\hat{\mathbf{b}}_k[u]^H\mathbf{A}\hat{\mathbf{h}}_k[u]\right]\right|^2 \tag{B.26}$$

$$\approx \left|\rho_d\sqrt{N}G\mathbf{E}\left[\hat{\mathbf{b}}_k[u]^H\hat{\mathbf{h}}_k[u]\right]\right|^2 \tag{B.27}$$

$$= |\rho_d\sqrt{N}G|^2 = \rho_d^2 NG^2, \tag{B.28}$$

where the approximation in (B.27) is due to the approximation in (B.25), and (B.28) is owing to ZF combining. What remains is to find the denominator terms of $\gamma_k[u]'$ as follows:

$$\text{Var}[g_k[u]'] + \text{Var}[w_k[u]] = \text{Var}[g_k[u]'] + \text{Var}[i_k[u]] + \text{Var}[n_k[u]] + \text{Var}[q_k[u]], \tag{B.29}$$

$$\begin{aligned}
\text{Var}[n_k[u]] &= \mathbf{E}[|\hat{\mathbf{b}}_k[u]^H\mathbf{A}\tilde{\mathbf{w}}[u]|^2] \\
&\approx G^2\mathbf{E}\left[\hat{\mathbf{b}}_k[u]^H\tilde{\mathbf{w}}[u]\tilde{\mathbf{w}}[u]^H\hat{\mathbf{b}}_k[u]\right] \\
&= G^2\mathbf{E}\left[\text{Tr}\left[\tilde{\mathbf{w}}[u]\tilde{\mathbf{w}}[u]^H\hat{\mathbf{b}}_k[u]\hat{\mathbf{b}}_k[u]^H\right]\right] \\
&= G^2\text{Tr}\left[\mathbf{E}\left[\tilde{\mathbf{w}}[u]\tilde{\mathbf{w}}[u]^H\right]\mathbf{E}\left[\hat{\mathbf{b}}_k[u]\hat{\mathbf{b}}_k[u]^H\right]\right] \\
&= G^2NN_o\mathbf{E}\left[\text{Tr}\left[\hat{\mathbf{b}}_k[u]\hat{\mathbf{b}}_k[u]^H\right]\right] \\
&= G^2NN_o\mathbf{E}\left[||\hat{\mathbf{b}}_k[u]||^2\right] \approx \frac{G^2NN_o}{(M-K)(1-\sigma_e^2)},
\end{aligned} \tag{B.30}$$

where in the last step, the approximation $\mathbf{E}\left[||\hat{\mathbf{b}}_k[u]||^2\right] \approx 1/\left((M-K)(1-\sigma_e^2)\right)$, is used, whose proof can be found in [2]. $\text{Var}[q_k[u]]$ can be found similarly as

$$\text{Var}[q_k[u]|\tilde{s}_k[u]] \approx N(2-4\pi)/\left((M-K)\left(1-\sigma_e^2\right)\right), \tag{B.31}$$

under the approximation $\mathbf{C}_{\mathbf{q}[u]} \approx N(2-4\pi)\mathbf{I}$, which is accurate under two conditions. The first is when L is large, for which the approximation $\mathbf{C}_{\mathbf{r}}[0] \approx G\mathbf{I}$ has been shown to be accurate, which along with (6.18) and (6.19) implies that $\mathbf{C}_{\mathbf{q}}[0] \approx (2-4/\pi)\mathbf{I}$ is accurate. The second condition is when the oversampling rates are low and when $\rho_d^2 \approx \rho_i^2$, in which case it can be shown that the approximation $\mathbf{C}_{\mathbf{r}}[\ell] \approx \mathbf{0}$ for $\ell \neq 0$ is accurate, which implies along with (6.18) and (6.19) that the approximation $\mathbf{C}_{\mathbf{q}}[\ell] \approx \mathbf{0}$ for $\ell \neq 0$ is accurate. Then, it follows from Proposition 3 that the approximation $\mathbf{C}_{\mathbf{q}[u]} \approx N(2-4/\pi)$ is accurate. For the proof of the approximation $\mathbf{C}_{\mathbf{r}}[\ell] \approx \mathbf{0}$ for

$\ell \neq 0$ being accurate for low oversampling rates, consider the received signal vector at the mth antenna, namely $\mathbf{r}_m \triangleq [r_m[0]\ r_m[1]\ \ldots r_m[N-1]]^T$. It can be written as $\mathbf{r}_m = \underline{\mathbf{H}}_m \underline{\mathbf{s}} + \underline{\mathbf{w}}_\mathbf{m}$, where $\underline{\mathbf{H}}_m$ is a block circulant channel matrix whose u^{th} row is equal to the $(mu)^{th}$ row of $\underline{\mathbf{H}}$ and $\underline{\mathbf{w}}_\mathbf{m}$ is a vector whose u^{th} element is equal to the $(mu)^{th}$ element of $\underline{\mathbf{w}}$. Moreover, defining $\underline{\tilde{\mathbf{s}}} \triangleq (\mathbf{F}_N \otimes \mathbf{I}_{K+I})\tilde{\mathbf{s}}$, where $\otimes$ represents the Kronecker product, $\mathbf{F}_N^H$ is the $N \times N$ DFT matrix, the autocorrelation of $\mathbf{r}_m$ conditioned on $\underline{\mathbf{H}}_\mathbf{m}$, namely $\mathbf{C}_{\mathbf{r}_m} \triangleq \mathbf{E}[\mathbf{r}_m \mathbf{r}_m^H | \underline{\mathbf{H}}_\mathbf{m}]$, can be found as

$$\mathbf{C}_{\mathbf{r}_m} = \underline{\mathbf{H}}_m \mathbf{P} \underline{\mathbf{H}}_m^H + N_o \mathbf{I}, \tag{B.32}$$

where $\mathbf{P} \triangleq \underline{\mathbf{F}}^H \mathbf{E}[\underline{\tilde{\mathbf{s}}}\underline{\tilde{\mathbf{s}}}^H]\underline{\mathbf{F}}$ and $\underline{\mathbf{F}}^H = (\mathbf{F}_N^H \otimes \mathbf{I}_{K+I})$. When $\rho_d^2 \approx \rho_i^2 = \rho^2$, the magnitude of the element of matrix $\underline{\mathbf{P}}$ at its mth row and n^{th} column, denoted by $|\underline{\mathbf{P}}_{m,n}|$, can be written as

$$|\underline{\mathbf{P}}_{m,n}| = \left| \frac{\rho^2}{N} \sum_{u=0}^{N-1} \mathrm{E}\left[|\tilde{s}_k[u]|^2\right] e^{-j2\pi\ell u/N} \right| \tag{B.33}$$

$$\overset{(*)}{=} \left| -\frac{\rho^2}{N} \sum_{u \notin (\mathcal{U}_D \cup \mathcal{U}_I)} \mathrm{E}\left[|\tilde{s}_k[u]|^2\right] e^{-j2\pi\ell u/N} \right| \tag{B.34}$$

$$< \frac{\rho^2}{N}(N - |\mathcal{U}_D| - |\mathcal{U}_I|), \tag{B.35}$$

where k is the user index determined by the value of m and $\ell = m - n$. Moreover, the equality (*) holds when $\ell \neq 0$. Since m^{th} diagonal element of $\underline{\mathbf{P}}$, namely $\underline{\mathbf{P}}_{m,m} = \rho^2/N\ \forall m$, the ratio of the magnitude of any non-diagonal element of $\underline{\mathbf{P}}$ to any diagonal element is bounded by $(N - |\mathcal{U}_D| - |\mathcal{U}_I|)$. Therefore, as $|\mathcal{U}_D| + |\mathcal{U}_I|$ approaches N, which occurs in a low oversampling rate scenario, the error in approximating $\underline{\mathbf{P}}$ as a diagonal matrix goes to zero. In this case, $\mathbf{C}_{\mathbf{r}_m}$, whose each element is a weighted summation of the channel coefficients $h_{m,k}[\ell]$, converges to $\mathrm{E}_{\underline{\mathbf{H}}_m}[\mathbf{C}_{\mathbf{r}_m}] = \mathrm{E}_{\underline{\mathbf{H}}_m}\left[\underline{\mathbf{H}}_m \underline{\mathbf{H}}_m^H\right] \underline{\mathbf{P}} + N_o \mathbf{I} = (K+I)\underline{\mathbf{P}} + N_o \mathbf{I}$ as L grows large, which can be shown rigorously following similar steps as in (B.22)-(B.24). Since the proof is the same $\forall m$, the error in approximating $\mathbf{C}_{\mathbf{r}_m}$ matrices as a diagonal matrices $\forall m$, equivalently approximating $\mathbf{C}_\mathbf{r}[\ell] \approx \mathbf{0}$ for $\ell \neq 0$ or $\mathbf{C}_{\mathbf{q}[u]} \approx N(2 - 4/\pi)\mathbf{I}$, goes to zero as L grows large and as $|\mathcal{U}_D| + |\mathcal{U}_I|$ approaches N (which is a case for low oversampling rates) when $\rho_d^2 \approx \rho_i^2$.

The proof continues with obtaining Var $[i_k[u]]$ as follows:

$$\text{Var}\left[i_k[u]\right] = \rho_d^2 N \text{Var}\left[\sum_{z\neq k, z\in\mathcal{K}_d} \hat{\mathbf{b}}_k[u]^H \mathbf{A}\hat{\mathbf{h}}_z[u]\tilde{s}_z[u] + \sum_{z\neq k, z\in\mathcal{K}_d} \hat{\mathbf{b}}_k[u]^H \mathbf{A}\tilde{\mathbf{e}}_z[u]\tilde{s}_z[u]\right] \tag{B.36}$$

$$\approx G^2\rho_d^2 N \sum_{z\neq k, z\in\mathcal{K}_d} \mathbf{E}\left[|\hat{\mathbf{b}}_k[u]^H\tilde{\mathbf{h}}_z[u]|^2\right] + \sum_{z\neq k, z\in\mathcal{K}_d} G^2\rho_d^2 N \mathbf{E}\left[|\tilde{\mathbf{e}}_z[u]^H\hat{\mathbf{b}}_k[u]|^2\right] \tag{B.37}$$

Here, $\mathbf{E}\left[|\hat{\mathbf{b}}_k[u]^H\tilde{\mathbf{h}}_z[u]|^2\right] = 0\ \forall z \neq k$ due to ZF combining and

$$\begin{aligned}\mathbf{E}\left[|\tilde{\mathbf{e}}_z[u]^H\hat{\mathbf{b}}_k[u]|^2\right] &= \mathbf{E}\left[\text{Tr}\left[\hat{\mathbf{b}}_k[u]\hat{\mathbf{b}}_k[u]^H\tilde{\mathbf{e}}_z[u]\tilde{\mathbf{e}}_z[u]^H\right]\right] \\ &= \text{Tr}\left[\mathbf{E}\left[\hat{\mathbf{b}}_k[u]\hat{\mathbf{b}}_k[u]^H\right]\mathbf{E}\left[\tilde{\mathbf{e}}_z[u]\tilde{\mathbf{e}}_z[u]^H\right]\right] \qquad \text{(B.38)} \\ &= \sigma_e^2\mathbf{E}\left[||\hat{\mathbf{b}}_k[u]||^2\right] \approx \frac{\sigma_e^2}{(M-K)(1-\sigma_e^2)}, \qquad \text{(B.39)}\end{aligned}$$

where (B.38) is due to the uncorrelatedness of the channel estimation error $\tilde{\mathbf{e}}_k[u]$ with the channel estimate $\hat{\mathbf{h}}_z[u]$, hence with $\hat{\mathbf{b}}_z[u]$. Using (B.37) and (B.39), it can be written that

$$\text{Var}\left[i_k[u]\right] = \text{Var}\left[g_k[u]' + i_k[u]\right] \approx \frac{G^2\rho_d^2 N(K-1)\sigma_e^2}{(M-K)(1-\sigma_e^2)}. \tag{B.40}$$

Similarly, Var $[g_k[u]']$ can be found as

$$\text{Var}\left[g_k[u]'\right] \approx \sigma_e^2(M-K)/(1-\sigma_e^2). \tag{B.41}$$

Then, from (B.29), (B.30), (B.31), (B.40) and (B.41) it follows that

$$\begin{aligned}&\text{Var}[g_k[u]'] + \text{Var}[\hat{i}_k[u] + \hat{n}_k[u] + \hat{q}_k[u]] \\ &\qquad \approx \frac{G^2NN_o + N(2-4/\pi) + G^2\rho_d^2 NK\sigma_e^2}{(M-K)(1-\sigma_e^2)},\end{aligned} \tag{B.42}$$

which implies the proposition statement along with (B.28). ■

APPENDIX C

PROOFS IN CHAPTER 7

C.1 Derivation of $\underline{\mathbf{A}}^{(p,1)}$, $\underline{\mathbf{A}}^{(p,m)}$, $\mathbf{C}^{1}_{\underline{\mathbf{q}}^{(p)}}$, $\mathbf{C}^{m}_{\underline{\mathbf{q}}^{(p)}}$

For a one-bit quantizer, assuming zero-mean Gaussian inputs[1], the following holds [19]:

$$\begin{aligned}\underline{\mathbf{A}}^{(p,1)} &= \sqrt{4/\pi}\mathrm{diag}(\mathbf{C}_{\underline{\mathbf{y}}^{(p)}})^{-0.5} \\ &= \sqrt{4/\pi}\mathrm{diag}\left(\left(\mathbf{X}\mathbf{X}^H \otimes \mathbf{I}_M\right) + N_o\mathbf{I}_{M\tau}\right)^{-0.5} \\ &= \mathbf{B}_1 \otimes \mathbf{I}_M, \end{aligned} \tag{C.1}$$

where $\mathbf{B}_1 = \sqrt{4/\pi}\mathrm{diag}(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau)^{-0.5}$. For multi-bit midrise uniform quantizers with Gaussian inputs[1], $\underline{\mathbf{A}}^{(p,m)}$ can be obtained as [19],

$$\begin{aligned}\underline{\mathbf{A}}^{(p,m)} &= \frac{\Delta}{\sqrt{\pi}}\mathrm{diag}(\mathbf{C}_{\underline{\mathbf{y}}^{(p)}})^{-0.5} \\ &\quad \times \sum_{i=1}^{2^q-1} \exp\left(-\Delta^2\left(i-2^{q-1}\right)^2 \mathrm{diag}\left(\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}\right)^{-0.5}\right) \\ &= \mathbf{B}_m \otimes \mathbf{I}_M, \end{aligned} \tag{C.2}$$

where Δ is the quantizer step size and $\mathbf{B}_m$ can be found as

$$\begin{aligned}\mathbf{B}_m =& \frac{\Delta}{\sqrt{\pi}}\mathrm{diag}\left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right)^{-0.5} \\ &\times \sum_{i=1}^{2^q-1} \exp\left(-\Delta^2\left(i-2^{q-1}\right)^2 \mathrm{diag}\left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right)^{-0.5}\right)\end{aligned}$$

[1] This approximation is accurate even when the transmitted symbols are not Gaussian. The input of the ADC at the mth antenna can be written as a sum of KL i.i.d. finite variance random variables. Owing to the central limit theorem (CLT), this summation converge to Gaussian as KL grows large. According to the Berry-Essen inequality, the difference between the standard normal cumulative distribution function (CDF) and the CDF of the signal (excluding the thermal noise part) at quantizer input (normalized such that it becomes unit variance) is less than 0.037 if $KL > 100$. Since there is also the Gaussian thermal noise, KL can be much less for an accurate Gaussian approximation (even $KL = 16$ is sufficient as shown in [114, Fig.4]).

The auto-covariance matrix of the distortion term at the quantizer output $\underline{\mathbf{q}}^{(p)}$ for one-bit quantizer case, namely $\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}}$, can be expressed using (7.8) as

$$\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}} = \mathbf{C}^1_{\underline{\mathbf{r}}^{(p)}} - \underline{\mathbf{A}}^{(p,1)}\mathbf{C}_{\underline{\mathbf{y}}^{(p)}}\left(\underline{\mathbf{A}}^{(p,1)}\right)^H, \tag{C.3}$$

where $\mathbf{C}^1_{\underline{\mathbf{r}}^{(p)}}$ can be found using *arcsine law* [64] as

$$\begin{aligned}\mathbf{C}^1_{\underline{\mathbf{r}}^{(p)}} = \frac{4}{\pi}\Big(&\operatorname{asin}\left(\mathbf{D}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\operatorname{Re}\{\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}}\}\mathbf{D}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\right)\\ &+ j\operatorname{asin}\left(\mathbf{D}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\operatorname{Im}\{\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}}\}\mathbf{D}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\right)\Big),\end{aligned} \tag{C.4}$$

where $\mathbf{D}_{\underline{\mathbf{y}}^{(p)}} = \operatorname{diag}\left(\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}}\right)$. Plugging $\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}} = \left(\mathbf{X}\mathbf{X}^H \otimes \mathbf{I}_M\right) + N_o\mathbf{I}_{M\tau} = \left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right) \otimes \mathbf{I}_M$ in (C.4), one can find that

$$\mathbf{C}^1_{\underline{\mathbf{r}}^{(p)}} = \mathbf{C}_\eta \otimes \mathbf{I}_M, \tag{C.5}$$

$$\begin{aligned}\mathbf{C}_\eta = \frac{4}{\pi}\Big(&\operatorname{asin}\left(\mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\operatorname{Re}\{\mathbf{G}_{\underline{\mathbf{y}}^{(\mathbf{p})}}\}\mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\right)\\ &+ j\operatorname{asin}\left(\mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\operatorname{Im}\{\mathbf{G}_{\underline{\mathbf{y}}^{(\mathbf{p})}}\}\mathbf{K}_{\underline{\mathbf{y}}^{(\mathbf{p})}}{}^{-\frac{1}{2}}\right)\Big),\end{aligned} \tag{C.6}$$

$\mathbf{G}_{\underline{\mathbf{y}}^{(p)}} = \left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right)$, $\mathbf{K}_{\underline{\mathbf{y}}^{(p)}} = \operatorname{diag}\left(\mathbf{G}_{\underline{\mathbf{y}}^{(p)}}\right)$. Then, it follows from (C.1), (C.3) and (C.5) that

$$\mathbf{C}^1_{\underline{\mathbf{q}}^{(p)}} = \mathbf{E}_1 \otimes \mathbf{I}_M, \tag{C.7}$$

where $\mathbf{E}_1 = \mathbf{C}_\eta - \mathbf{B}_1\left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right)\mathbf{B}_1^H$. For multi-bit quantizer case, $\mathbf{C}_{\underline{\mathbf{q}}^{(p)}} = \mathbf{C}^m_{\underline{\mathbf{q}}^{(p)}}$, where $\mathbf{C}^m_{\underline{\mathbf{q}}^{(p)}}$ can be approximated as a diagonal matrix as [19]

$$\begin{aligned}\mathbf{C}^m_{\underline{\mathbf{q}}^{(p)}} \approx&\frac{\Delta^2}{2}(2^q-1)^2\mathbf{I}_{M\tau} - \underline{\mathbf{A}}^{(p,m)}\operatorname{diag}\left(\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}}\right)(\underline{\mathbf{A}}^{(p,m)})^H\\ &- 4\Delta^2\sum_{i=1}^{2^q-1}\left(i-2^{q-1}\right)\times\left(1-\mathcal{Q}\left(\sqrt{2}(i-2^{q-1})\operatorname{diag}(\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}})^{-1/2}\right)\right)\\ =&\mathbf{E}_m \otimes \mathbf{I}_M,\end{aligned} \tag{C.8}$$

where $\mathbf{E}_m$ can be found by plugging $\mathbf{C}_{\underline{\mathbf{y}}^{(\mathbf{p})}} = \left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right) \otimes \mathbf{I}_M$ in (C.8) and using (C.2) as

$$\begin{aligned}\mathbf{E}_m \approx&\frac{\Delta^2}{2}(2^q-1)^2\mathbf{I}_\tau - \mathbf{B_m}\operatorname{diag}\left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right)\mathbf{B_m}{}^H\\ &- 4\Delta^2\sum_{i=1}^{2^q-1}\left(i-2^{q-1}\right)\times\left(1-\mathcal{Q}\left(\sqrt{2}(i-2^{q-1})\operatorname{diag}\left(\mathbf{X}\mathbf{X}^H + N_o\mathbf{I}_\tau\right)^{-1/2}\right)\right).\end{aligned} \tag{C.9}$$

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Üçüncü, Ali Bulut

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Middle East Technical University	2015
B.S.	Middle East Technical University	2012

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
09/2012 -	Middle East Technical University	Research/Teaching Assistant

PUBLICATIONS

1) A. B. Üçüncü, G. M. Güvensen, and A. Ö. Yılmaz, "A Reduced Complexity Ungerboeck Receiver for Quantized Wideband Massive SC-MIMO." in IEEE Transactions on Communications, vol. 69, no. 7, pp. 4921-4936, Jul. 2021.

2) A. B. Üçüncü, E. Björnson, H. Johansson, A. Ö. Yılmaz and E. G. Larsson, "Performance Analysis of Quantized Uplink Massive MIMO-OFDM With Oversampling Under Adjacent Channel Interference," in IEEE Transactions on Communications, vol. 68, no. 2, pp. 871-886, Feb. 2020.

3) A. B. Üçüncü and A. Ö. Yılmaz, “Oversampling in one-bit quantized massive

MIMO systems and performance analysis," IEEE Transactions on Wireless Communications, vol. 17, no. 12, pp. 7952–7964, Dec. 2018.

INTERNATIONAL PATENTS

1) A. Ö. Yılmaz, A. B. Üçüncü, "Quantized detection in uplink MIMO with oversampling", US Patent 10,447,504, Oct. 15, 2019.

International Conference Publications

1) A. B. Üçüncü, E. Björnson, H. Johansson, A.Ö. Yılmaz and E.G. Larsson, "Performance analysis of one-bit massive MIMO with oversampling under adjacent channel interference," in Proc. IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 2019, pp. 1-6,

2) A. B. Üçüncü and A. Ö. Yılmaz, "Uplink performance analysis of oversampled wideband massive MIMO with one-bit ADCs," in Proc. IEEE 88th Vehicular Technology Conference, 2018, pp. 1–5.

3) A. B. Üçüncü and A. Ö. Yılmaz, "Sequential linear detection in one-bit quantized uplink massive MIMO with oversampling," in Proc. IEEE 88th Vehicular Technology Conference, 2018, pp. 1–5.

4) A. B. Üçüncü and A. Ö. Yılmaz, "Performance analysis of faster than symbol rate sampling in 1-bit massive MIMO systems," in Proc. IEEE International Conference on Communications, 2017, pp. 1–6.

National Conference Publications

1) A. B. Üçüncü and A. Ö. Yilmaz, "A new sampling method for massive MIMO systems," in Proc. 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.

2) A. B. Üçüncü and A. Ö. Yılmaz, "Comparison of new multi-carrier modulation

schemes under amplifier nonlinearity," in Proc. 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, 2016, pp. 249-252.